DOCUMENT RESUME

ED 342 795 TM 017 957

AUTHOR Groenewald, A. C.; Stoker, D. J.
TITLE Estimation of Variance in the Case of Complex Samples.
INSTITUTION Human Sciences Research Council, Pretoria (South Africa).
REPORT NO HSRC-WS-43; ISBN-0-7969-0836-2
PUB DATE 90
NOTE 123p.; M.S. Thesis, University of South Africa. An Afrikaans abstract and an English abstract are included. A product of the Institute for Statistical Research.
PUB TYPE Dissertations/Theses - Masters Theses (042) -- Reports - Evaluative/Feasibility (142)

EDRS PRICE MF01/PC05 Plus Postage.
DESCRIPTORS *Equations (Mathematics); *Estimation (Mathematics); Foreign Countries; *Mathematical Models; Probability; *Research Design; Sample Size; *Sampling; Surveys
IDENTIFIERS Population Parameters; *Variance (Statistical)

ABSTRACT

In a complex sampling scheme it is desirable to select the primary sampling units (PSUs) without replacement to prevent duplications in the sample. Since the estimation of the sampling variances is more complicated when the PSUs are selected without replacement, L. Kish (1965) recommends that the variance be calculated using the formulas applicable for sampling with replacement. Kish states that this method probably leads to an overestimation of the true variance. This assertion, which is not self-evident, is investigated in this study. If the PSUs are selected in such a way that PSU "i" has a probability pi(sub i) equals nz(sub i) of being included in the sample, the population total is estimated unbiasedly. This is not the case if pi(sub i) is not equal to nz(sub i). For these two cases, various sampling procedures are considered. It is numerically verified that this assertion is true on average regardless of whether or not pi(sub i) equals nz(sub i). A sampling procedure for practical applications is recommended. There are 28 tables in the text and an additional 6 tables in 2 appendices. There is also a 25-item list of references. (Author/SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

TJAART VAN DER WALT

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

ED342795

INSTITUTE FOR STATISTICAL RESEARCH WS-43

Estimation of variance in the case of complex samples

A.C. Groenewald
D.J. Stoker

RGN·HSRC

TM 017 957

BEST COPY AVAILABLE

Estimation of variance in the case of complex samples

Report WS-43 1990

Estimation of variance in the case of complex samples

A.C. Groenewald
D.J. Stoker

Pretoria
Human Sciences Research Council
1990

A.C. Groenewald, M.Sc.
D.J. Stoker, Math. et Phys. Dr.

Institute for Statistical Research
Executive Director: N.A.S. Crowther

ISBN 0 7969 0836 2

© Human Sciences Research Council, 1990

Printed by HSRC Publishers
134 Pretorius Street, Pretoria

PREFACE

When investigating a certain characteristic or variable by means of a sample survey, any inferential or hypothesis-testing procedure relies heavily on the variance of the underlying distribution of the variable. The estimation of parameters in various statistical models is also based on this variance. Consequently an accurate assessment or estimate of the variance in specific complex sampling schemes is of the utmost importance.

This report makes a contribution to the correct estimation of such variance. It is believed that the results of the investigation will lead to the sound and correct application of a variety of statistical procedures in complex sampling schemes.

This report has been submitted in part fulfilment of the requirements for the degree of Master of Sciences in Statistics at the University of South Africa and has been completed under the supervision of Prof. D.J. Stoker with Prof. M. Schultz as joint supervisor.

EKSERP

Dit is wenslik om in 'n komplekse steekproefnemingskema die primêre steekproefnemingseenhede (PSE's) sonder teruglegging te trek ten einde duplikate in die steekproef te voorkom. Aangesien die beraming van die steekproefvariansies meer gekompliseerd is as die PSE's sonder teruglegging getrek word, beveel Kish (1965) aan dat vir die variansieberekening die formules vir steekproefneming met teruglegging gebruik moet word. Hy beweer dat hierdie prosedure tot 'n matige oorberaming van die werklike variansie lei. Hierdie bewering, wat nie vanself volg nie, word in hierdie verslag ondersoek. As die PSE's op so 'n wyse getrek word dat die i-de PSE 'n waarskynlikheid $\pi_i = nz_i$ het om as element van die steekproef getrek te word, lei Kish se prosedure tot onsydige beraming van die populasietotaal. Dit geld egter nie in die geval $\pi_i \neq nz_i$ nie. Verskillende steekproefnemingsmetodes is vir hierdie twee gevalle numeries ondersoek en dit is bevind dat die bewering normaalweg geld ongeag of $\pi_i = nz_i$ al dan nie. 'n Steekproefnemingsprosedure word aanbeveel vir toepassing in die praktyk.

ABSTRACT

In a complex sampling scheme it is desirable to select the primary sampling units (PSUs) without replacement to prevent duplications in the sample. Since the estimation of the sampling variances is more complicated when the PSUs are selected without replacement, Kish (1965) recommends that the variance is calculated using the formulas applicable for sampling with replacement. He states that this method probably leads to an overestimation of the true variance. This assertion is not self-evident and is investigated in this study.

If the PSUs are selected in such a way that PSU i has probability $\pi_i = nz_i$ of being included in the sample, the population total is estimated unbiasedly. This is not the case if $\pi_i \neq nz_i$. For these two cases various sampling procedures are considered and it is numerically verified that this assertion is true on the average regardless whether $\pi_i = nz_i$ or not. A sampling procedure for practical applications is recommended.

TABLE OF CONTENTS

PAGE

CHAPTER 1 : AN INTRODUCTION INTO THE ESTIMATION OF VARIANCE IN COMPLEX SAMPLES

1.1 INTRODUCTION 1

1.2 COMPLEX SAMPLING 1

1.3 SOME BASIC ASPECTS 5

1.4 SAMPLING OF PSUs WITH REPLACEMENT 12

1.4.1 Single-stage cluster sampling 12

1.4.2 Multistage cluster sampling 14

1.5 SAMPLING OF PSUs WITHOUT REPLACEMENT 17

1.5.1 Single-stage cluster sampling 17

1.5.2 Multistage cluster sampling 21

1.6 THE PROBLEM 24

CHAPTER 2 : SAMPLING WITHOUT REPLACEMENT IN THE CASE WHERE $\pi_i = nz_i$

2.1 THE EFFECT OF AN INAPPROPRIATE VARIANCE ESTIMATOR 27

2.1.1 Introduction 27

2.1.2 Single-stage sampling 27

2.1.3 Multistage sampling 30

2.1.4 Conclusion 32

2.2 SAMPLING PROCEDURES FOR WHICH $\pi_i = nz_i$ 32

2.2.1 Introduction 32

2.2.2 Procedures suitable for $n = 2$ 33

2.2.3 Procedures suitable for $n > 2$ 37

PAGE

2.3 NUMERICAL INVESTIGATIONS 48

2.3.1 Introduction 48

2.3.2 An investigation into the degree of overestimation 50

2.3.3 Comparison between $\hat{Var}_{wr}(\hat{Y}_{pps})$ and $\hat{Var}_{wor}(\hat{Y}_{pps})$ 53

CHAPTER 3 : SAMPLING WITHOUT REPLACEMENT IN THE CASE WHERE $\pi_i \neq nz_i$

3.1 THE EFFECT OF AN INAPPRORIATE VARIANCE ESTIMATOR 59

3.2. SAMPLING PROCEDURES FOR WHICH $\pi_i \neq nz_i$ 60

3.2.2 Successive sampling 60

3.2.3 Rejective sampling 67

3.3 NUMERICAL INVESTIGATIONS 80

3.3.1 An investigation into the bias 81

3.3.2 An investigation into the over-/underestimation of $Var_{wor}(\hat{Y}_{pps})$ 85

3.3.3 An investigation into the over-/underestimation of $Var_{wor}(\hat{Y}_{HT})$ 88

3.3.4 Comparison between $\hat{Var}_{wr}(\hat{Y}_{pps})$ and $\hat{Var}_{wor}(\hat{Y}_{pps})$ 90

CHAPTER 4 : CONCLUSIONS AND RECOMMENDATIONS 98

APPENDIX A 100

APPENDIX B 105

REFERENCES 109

SUMMARY

In a complex sampling scheme it is desirable to select the primary sampling units (PSUs) without replacement to prevent duplications in the sample. Since the estimation of the sampling variances is more complicated when the PSUs are selected without replacement, Kish (1965) recommends that the variance is calculated using the formulas applicable for sampling with replacement. He states that this method probably leads to an overestimation of the true variance. This assertion is not self-evident and is investigated in this study.

If the PSUs are selected in such a way that PSU i has probability $\pi_i = nz_i$ of being included in the sample, the population total is estimated unbiasedly. This is not the case if $\pi_i \neq nz_i$. For these two cases various sampling procedures are considered and it is numerically verified that this assertion is true on the average regardless whether $\pi_i = nz_i$ or not. A sampling procedure for practical applications is recommended.

CHAPTER 1

AN INTRODUCTION INTO THE ESTIMATION OF VARIANCE IN COMPLEX SAMPLES

1.1 INTRODUCTION

In a complex sampling scheme, a sample of n primary sampling units (PSUs) is to be drawn from a population of N PSUs in the first sampling stage. This sampling of PSUs can either take place with or without replacement. When the PSUs are selected with replacement, simple formulas for the estimation of the variance of the population total estimator are available. However, sampling with replacement is often undesirable as it permits any PSU to appear in the sample more than once. On the other hand, if the PSUs are selected without replacement, a sample of n PSUs is obtained but the formulas for the calculation of the variance estimator of the population total estimator are complicated. In order to overcome this problem, it is often customary to select the PSUs without replacement but to calculate the variance using the formulas applicable for sampling with replacement . Kish (1965) states that this method probably leads to an overestimation of the true variance of the population total estimator. This however is not self-evident and will be investigated in this study.

In this first chapter, complex sampling, sampling of PSUs with and without replacement and the problems encountered with these sampling methods are discussed.

1.2 COMPLEX SAMPLING

A complex sample is a stratified multistage cluster sample.

According to Cochran (1977) and Stoker (1988), stratification is used for the following reasons:

(i) Good representation of the population in the sample can be ensured. The population can be stratified with regard to certain characteristics of the population (for example race or province) which are expected to have an influence on the population characteristic to be investigated and which must therefore be represented in the sample.

A simple random sample does not guarantee that the sample is representative of the population with regard to these important characteristics. Certain categories or values of certain characteristics may be overrepresented in the sample, while others may be underrepresented.

(ii) The precision of the estimate of a population parameter being investigated can be increased by using stratification. This will be the case when each stratum is more homogeneous with regard to the characteristic being investigated than the population as a whole. Thus, the variation in the values of the characteristic concerned will be less for the elements within the same stratum than that for the elements in the different strata. Stratification therefore leads to population parameter estimators with smaller standard deviations.

(iii) If it is desirable to obtain separate estimates of a population parameter for different parts or study domains of the population, strata containing these parts or study domains in their entirety can be formed and each study domain can be considered as a separate population.

(iv) Administrative convenience can also play a role in forming strata.

(v) Different parts of the population may have different sampling problems. By defining each part as a stratum, these problems may be dealt with separately.

Cluster sampling on the other hand is applied for the following reasons (Cochran (1977) and Stoker (1988)):

(i) In many surveys no reliable list of the population elements is available and it may be very expensive and time-consuming to compile one, which is a prerequisite if simple random sampling is to be used. With cluster sampling the population elements are grouped into clusters and several clusters are drawn. For example, if a sample of people in an urban area is to be drawn, a complete list of persons living in that area is often not readily available. A street map can be used, and clusters can be formed consisting of street blocks or by grouping together adjacent street blocks. Several clusters are then drawn and now it is only necessary to list the people living in the drawn clusters.

(ii) Cluster sampling is generally more economical and practical than a simple random sample. The elements in the sample are grouped together into clusters which implies that travelling costs are reduced and time is saved.

Cluster sampling may however result in a loss in the precision of the estimate of the population parameter being investigated. Generally geographically adjacent elements, which tend to be more homogeneous than elements lying further apart, are grouped together in a cluster. Thus the elements in a cluster are more homogeneous with regard to the

characteristic being investigated than the elements of the entire population, which results in a larger standard error.

If the population is very large it is usually preferable to sample in several stages as this is more economical and practical.

The general complex sampling procedure is as follows (Kish (1965) and Stoker (1988)):

(i) Firstly, the population is divided into several strata. In order to achieve a reduction in the variance of the estimator of the population parameter under investigation, it is advisable that each stratum formed is more homogeneous with regard to the characteristic being investigated, than the entire population.
(ii) In the first sampling stage the population elements in each stratum are grouped into a number of meaningful clusters (units). These newly formed clusters are called primary sampling units (PSUs). A predetermined number of these PSUs are drawn by some or other probability selection procedure.
(iii) In the second sampling stage the population elements in each selected PSU are again grouped into smaller meaningful clusters, called secondary sampling units (SSUs). Again a predetermined number of SSUs are drawn from each selected PSU.
(iv) This process can be continued in a similar way until eventually population elements or groups of population elements, called ultimate sampling units (USEs), are drawn.

In most surveys it is desirable that the sample should be self-weighting. This means that every element in the population has the same probability of being selected. Stoker (1984) states the following advantages of a self-weighting sample. Firstly, the calculation of the sample mean or sample proportion is easier, as the data need not be weighted. Secondly, the variance of the sample mean or proportion is usually less if the data originates from a self-weighting sample than from a non-self-weighting sample.

In a two-stage cluster sample, for example, this self-weighting property can be achieved in two ways:

(i) The PSUs in the first sampling stage are selected with equal probabilities whereafter a fixed proportion of elements is selected from each selected PSU. However, if the fixed proportion is determined in advance, the sample size is not fixed; and
(ii) The PSUs in the first sampling stage are selected with probabilities proportional to the number of elements contained in each PSU whereafter a fixed number of elements is selected from each selected PSU.

In this study the estimation of the population total will be considered. It is supposed that the population consists of only one stratum. In the case where the population consists of several strata, the population total estimator is obtained by simply adding the estimates of the population total obtained for each stratum, over all the strata. The same is true for the variance estimator. This can be done since sampling in each stratum takes place independently.

Suppose that the population comprises N primary units of which n have to be selected in successive draws. Though many types of selection procedures exist, only probability sampling will be considered here. In probability sampling, the sampling procedure will yield a certain sample s of size n with probability $P(s)$. The probabilities $P(s)$ satisfy

$$\sum_{s} P(s) = 1 \tag{1.1}$$

where summation $\sum_s$ extends over all possible samples of size n.

Many types of probability sampling procedures do exist.

The population total of a certain characteristic or variable is defined as

$$Y = \sum_{i=1}^{N} y_i \tag{1.2}$$

where y_i is the total of the characteristic y for the i-th primary unit.

For each draw, Y can be estimated unbiasedly by

$$(\hat{Y})_i = \frac{y_i}{p_i} \tag{1.3}$$

where p_i is the probability of drawing PSU i on that draw.

The probability of drawing PSU i on a specific draw may differ from sampling scheme to sampling scheme. Thus, $(\hat{Y})_i$ depends on the sampling scheme applied. To estimate Y after n draws depends on whether sampling with or without replacement has been applied. Sampling of PSUs with replacement will be discussed in Section 1.4 and sampling of PSUs without replacement in Section 1.5. In each case single-stage sampling will be considered first and then extended to multistage sampling.

Also note that some or other measure of size (MOS) is associated with each primary unit. This measure of size could be some historical data which provides information on the parameter under investigation, such as a previous sensus; or it could be some other characteristics of the PSUs, such as the number of people, farms, or households comprising these units. The MOS of the i-th PSU is denoted by M_i, $i = 1, 2, ..., N$, where

$$M_0 = \sum_{i=1}^{N} M_i$$

is the total measure of size of the N primary units contained in the population.
It is often customary to select the sample of n PSUs with probability proportional to these MOS to utilise prior information on the population to increase the precision of sampling. This will become apparent later. Then, each PSU could have a different probability of being included in the sample.

In this study we will always suppose that there are N PSUs in the population of which n are to be selected in the first sampling stage.

1.3 SOME BASIC ASPECTS

Note that the notation and formulas used in this paragraph are due to Horvitz and Thompson (1952), Konijn (1973), Jessen (1978) and Hajek (1981).

Consider a population of N PSUs from which a sample of size n is to be selected.
Define the probability of selecting primary unit i prior to the first draw as the *initial selection probability* of PSU i and let it be denoted by z_i, $i = 1, 2, ..., N$. The selection probabilities, z_i, can be any set of probabilities, i.e.

$$z_i \geq 0 \quad \text{and} \quad \sum_{i=1}^{N} z_i = 1 \quad .$$

It is often customary to take z_i as the relative MOS of PSU i, i.e. $z_i = M_i/M_0$.

When sampling with replacement is applied, the probability of selecting PSU i is the same for each draw.
If sampling takes place without replacement, prior to each succeeding draw a new set of selection probabilities for the remaining PSUs has to be defined. To distinguish these probabilities from the selection probabilities, z_i, they will be called *drawing probabilities*. The drawing probabilities may be based on the initial selection probabilities, but this is not a necessity. Denote the drawing probabilities for the m-th draw by p_{i_m}, $i = 1, 2, ..., N$, so that

$$p_{i_m} \geq 0 \quad \text{and} \quad \sum_{i} p_{i_m} = 1 \quad ,$$

where $m = 1, 2, ..., n$ and the summation $\sum_i$ extends over the remaining $N - m + 1$ primary units in the population prior to the m-th draw.

Also, as the n selections take place in successive steps, define $\delta_i(k)$ as the *k-draw selection probability* of primary unit i, which is the probability of selecting primary unit i in the k-th draw. For sampling with replacement, $\delta_i(k) = z_i$, $i = 1, 2, ..., N$ and $k = 1, 2, \ldots, n$.
Similarly, define $\delta_{ij}(k, l)$ as the *k-and-l-draw selection probability* of primary unit i and primary unit j, which is the probability of selecting PSU i at the k-th draw and PSU j in the l-th draw.

Define S as the number of possible samples of size n if the order of draw is taken into account. For sampling with replacement, $S = N^n$ and for sampling without replacement, $S = \binom{N}{n} n!$, provided that for each draw the remaining PSUs have probabilities greater than zero of being selected. Now define s_n as the s-th of these S samples of size n.

The probability that a sample s_n will be drawn is thus given by

$$P(s_n) = z_i z_j z_k \ldots z_t \tag{1.4}$$

for sampling with replacement, where $z_i z_j z_k \ldots z_t$ contains n terms, and by

$$P(s_n) = p_{i_1} p_{j_2} p_{k_3} \cdots p_{t_n} \tag{1.5}$$

for sampling without replacement if the sample contains the primary units $i, j, k, \ldots, t$, drawn in this order.

Let $s_n \ni i$ denote the event that sample s_n contains PSU i. Then, if $I_i(s_n)$ is defined by

$$I_i(s_n) = \begin{cases} 1 & \text{if} \quad s_n \ni i \\ 0 & \text{otherwise} \end{cases} \quad , \qquad 1 \le i \le N \quad , \tag{1.6}$$

$I_i(s_n)$ is the indicator of inclusion of PSU i in the sample s_n. The probability that primary unit i appears in a sample of size n is called the *inclusion probability* of primary unit i and is defined by

$$\pi_i = E[I_i(s_n)] = \sum_{s_n \ni i} P(s_n) \quad , \qquad 1 \le i \le N \quad , \tag{1.7}$$

where summation $\sum_{s_n \ni i}$ extends over all possible samples s_n that contain primary unit i.

Firstly consider sampling with replacement. In this case, a certain PSU may appear in the sample more than once. Thus, $1 - \pi_i$ is the probability that primary unit i does not appear in the sample. Since the n drawings take place independently,

$$1 - \pi_i = (1 - z_i)^n \quad .$$

Thus,

$$\begin{aligned} \pi_i &= 1 - (1 - z_i)^n \\ &= n z_i - \binom{n}{2} z_i^2 + \binom{n}{3} z_i^3 + \ldots + (-1)^{n-1} \binom{n}{n} z_i^n \quad . \end{aligned} \tag{1.8}$$

For sampling without replacement, the inclusion probability is also defined by

$$\pi_i = \sum_{k=1}^{n} \delta_i(k) \quad . \tag{1.9}$$

The extension to the probability of including both primary units i and j in the sample follows readily. The *joint inclusion probability*, π_{ij}, of both primary units i and j in the sample is defined by

$$\pi_{ij} = \sum_{s_n \ni i,j} P(s_n) \quad , \quad 1 \leq i \neq j \leq N \quad , \tag{1.10}$$

where summation $\sum_{s_n \ni i,j}$ extends over all the different samples, s_n, that contain both primary units i and j.

For sampling without replacement, the joint inclusion probability is also defined by

$$\pi_{ij} = \sum_{k=1}^{n} \sum_{l \neq k}^{n} \delta_{ij}(k,l) \quad . \tag{1.11}$$

Further, for sampling without replacement, define $\overline{\pi}_i$ as the *average probability* of primary unit i being drawn on one of the n draws. If the order of draw is ignored, $\overline{\pi}_i$ is the probability associated with PSU i to be drawn on each of the n draws. This average probability is defined by

$$\overline{\pi}_i = \frac{\pi_i}{n} \quad , \quad i \leq 1 \leq N \quad . \tag{1.12}$$

The following useful relations hold for sampling without replacement, provided the sample size, n, is fixed (Cochran (1977)):

$$\sum_{i=1}^{N} \pi_i = n \quad , \quad \sum_{j \neq i}^{N} \pi_{ij} = (n-1)\pi_i \quad \text{and} \quad \sum_{i=1}^{N} \sum_{j>i}^{N} \pi_{ij} = \frac{1}{2} n(n-1) \quad . \tag{1.13}$$

Also, according to Hajek (1981) and Konijn (1973),

$$\sum_{j \neq i}^{N} (\pi_i \pi_j - \pi_{ij}) = \pi_i(1 - \pi_i) \quad . \tag{1.14}$$

Note that these relations do not hold for sampling with replacement.

The following two examples will illustrate the foregoing theory. For these examples, consider the following universe from which a sample of size two is to be drawn:

PSU, i	1	2	3	4
Measure of size, M_i	10	20	15	5
Relative measure of size, M_i/M_0	0,2	0,4	0,3	0,1

where $M_0 = \sum_{i=1}^{N} M_i = 50$.

Suppose that PSU i is drawn first and PSU j second.

a) Example 1: Sampling with replacement

The two PSUs are selected with replacement with selection probabilities proportional to the measures of size. The following $4^2 = 16$ samples s_n are possible:

Sample no.	*Sample*	*Selection Probabilities*		$P(s_n)$
	$s_n = (i, j)$	z_i	z_j	$z_i.z_j$
1	1, 1	0, 2	0, 2	0, 04
2	1, 2	0, 2	0, 4	0, 08
3	1, 3	0, 2	0, 3	0, 06
4	1, 4	0, 2	0, 1	0, 02
5	2, 1	0, 4	0, 2	0, 08
6	2, 2	0, 4	0, 4	0, 16
7	2, 3	0, 4	0, 3	0, 12
8	2, 4	0, 4	0, 1	0, 04
9	3, 1	0, 3	0, 2	0, 06
10	3, 2	0, 3	0, 4	0, 12
11	3, 3	0, 3	0, 3	0, 09
12	3, 4	0, 3	0, 1	0, 03
13	4, 1	0, 1	0, 2	0, 02
14	4, 2	0, 1	0, 4	0, 04
15	4, 3	0, 1	0, 3	0, 03
16	4, 4	0, 1	0, 1	0, 01
			Total	1, 00

From (1.7) it can be seen that the inclusion probability, π_i, is the sum of $P(s_n)$ over all samples, s_n, containing PSU i. For instance,

$$\begin{aligned}\pi_1 &= \sum_{s_n \ni 1} P(s_n) \\ &= P(1,1) + P(1,2) + P(1,3) + P(1,4) + P(2,1) + P(3,1) + P(4,1) \\ &= 0,04 + 0,08 + 0,06 + 0,02 + 0,08 + 0,06 + 0,02 \\ &= 0,36.\end{aligned}$$

The following inclusion probabilities are obtained:

$$\pi_1 = 0,36 \quad , \quad \pi_2 = 0,64 \quad , \quad \pi_3 = 0,51 \quad , \quad \pi_4 = 0,19 \quad .$$

Thus,

$$\sum_{i=1}^{4} \pi_i = 1,70 \neq 2.$$

π_i can be readily obtained by using (1.8). For instance,

$$\begin{aligned}\pi_1 &= n z_1 - \binom{n}{2} {z_1}^2 \\ &= 0,4 - (0,2)^2 \\ &= 0,36.\end{aligned}$$

The joint inclusion probability, π_{ij}, is defined by (1.10). For instance,

$$\begin{aligned}\pi_{12} &= \sum_{s_n \ni 1,2} P(s_n) \\ &= P(1,2) + P(2,1) \\ &= 0,08 + 0,08 \\ &= 0,16.\end{aligned}$$

The following joint inclusion probabilities are obtained:

$$\pi_{12} = 0,16 \quad , \quad \pi_{13} = 0,12 \quad , \quad \pi_{14} = 0,04 \quad ,$$
$$\pi_{23} = 0,24 \quad , \quad \pi_{24} = 0,08 \quad , \quad \pi_{34} = 0,06 \quad .$$

Thus,

$$\sum_{i=1}^{4} \sum_{j>i}^{4} \pi_{ij} = 0,7 \neq \frac{1}{2} n(n-1) = 1.$$

b) Example 2: Sampling without replacement

The two PSUs are selected without replacement. At each draw, one PSU is selected from the remaining PSUs with probability exactly proportional to the measures of size, where the relative measures of size are modified after each draw. Thus, the first PSU is selected with drawing probability exactly equal to the initial selection probability, i.e. $p_{i_1} = z_i$. On the second draw PSU j is selected with drawing probability $p_{j_2} = z_j/(1-z_i)$. The following $\binom{4}{2} 2! = 12$ samples s_n are possible:

Sample no.	*Sample*	*Drawing Probabilities*		$P(s_n)$
	$s_n = (i,j)$	p_{i_1}	p_{j_2}	$p_{i_1} \cdot p_{j_2}$
1	1, 2	0, 2	0, 5000	0, 1000
2	1, 3	0, 2	0, 3750	0, 0750
3	1, 4	0, 2	0, 1250	0, 0250
4	2, 1	0, 4	0, 3333	0, 1333
5	2, 3	0, 4	0, 5000	0, 2000
6	2, 4	0, 4	0, 1667	0, 0667
7	3, 1	0, 3	0, 2857	0, 0857
8	3, 2	0, 3	0, 5714	0, 1714
9	3, 4	0, 3	0, 1429	0, 0429
10	4, 1	0, 1	0, 2222	0, 0222
11	4, 2	0, 1	0, 4444	0, 0444
12	4, 3	0, 1	0, 3333	0, 0333
			Total	1, 0000

From (1.7) it can be seen that the inclusion probability, π_i, is the sum of $P(s_n)$ over all samples, s_n, containing primary unit i. For instance,

$$\begin{aligned}\pi_1 &= \sum_{s_n \ni 1} P(s_n) \\ &= P(1,2) + P(1,3) + P(1,4) + P(2,1) + P(3,1) + P(4,1) \\ &= 0,1000 + 0,0750 + 0,0250 + 0,1333 + 0,0857 + 0,0222 \\ &= 0,4412.\end{aligned}$$

The following inclusion probabilities are obtained:

$$\pi_1 = 0,4412 \quad , \quad \pi_2 = 0,7159 \quad , \quad \pi_3 = 0,6083 \quad , \quad \pi_4 = 0,2346 \quad .$$

Thus,

$$\sum_{i=1}^{4} \pi_i = 2 = n \quad .$$

Also, π_i can be obtained by using (1.9), resulting in the following formula:

$$\begin{aligned}\pi_i &= \delta_i(1) + \delta_i(2) \\ &= P(\text{PSU } i \text{ drawn first}) + P(\text{PSU } i \text{ drawn second}) \\ &= z_i + \sum_{j \neq i}^{N} z_j \frac{z_i}{1 - z_j} \\ &= z_i \left(1 + \sum_{j \neq i}^{N} \frac{z_j}{1 - z_j} \right) \\ &= z_i \left(1 + \sum_{j=1}^{N} \frac{z_j}{1 - z_j} - \frac{z_i}{1 - z_i} \right) \quad .\end{aligned}$$

For instance, π_1 is given by

$$\begin{aligned}\pi_1 =& 0,2\left(1+\sum_{j=1}^{4}\frac{z_j}{1-z_j}-\frac{0,2}{0,8}\right)\\ =& 0,2(1+1,4563-0,25)\\ =& 0,4412.\end{aligned}$$

From (1.10) it can be seen that the joint inclusion probability, π_{ij}, is the sum of $P(s_n)$ over all samples, s_n, containing primary units i and j. For instance,

$$\begin{aligned}\pi_{12} =& \sum_{s_n \ni 1,2} P(s_n)\\ =& P(1,2)+P(2,1)\\ =& 0,1000+0,1333\\ =& 0,2333.\end{aligned}$$

The following joint inclusion probabilities are obtained:

$$\pi_{12}=0,2333 \quad , \quad \pi_{13}=0,1607 \quad , \quad \pi_{14}=0,0472 \quad ,$$
$$\pi_{23}=0,3714 \quad , \quad \pi_{24}=0,1112 \quad , \quad \pi_{34}=0,0762 \quad .$$

Thus,

$$\sum_{i=1}^{4}\sum_{j>i}^{4}\pi_{ij}=1=\frac{1}{2}n(n-1).$$

Also, π_{ij} can be obtained by using (1.11), resulting in the following formula:

$$\begin{aligned}\pi_{ij} =& \delta_{ij}(1,2)+\delta_{ij}(2,1)\\ =& P(\text{PSU } i \text{ drawn first, PSU } j \text{ drawn second})\\ &+P(\text{PSU } i \text{ drawn second, PSU } j \text{ drawn first})\\ =& z_i\frac{z_j}{1-z_i}+z_j\frac{z_i}{1-z_j}\\ =& z_iz_j\left(\frac{1}{1-z_i}+\frac{1}{1-z_j}\right) \quad .\end{aligned}$$

For instance, π_{12} is given by

$$\begin{aligned}\pi_{12} =& (0,2)(0,4)(\frac{1}{0,8}+\frac{1}{0,6})\\ =& 0,2333.\end{aligned}$$

1.4 SAMPLING OF PSUs WITH REPLACEMENT

Sampling with replacement where the PSUs are selected with constant selection probabilities in each draw, is also known as multinomial sampling.
For a sample of size n, one PSU is selected at each draw, regardless whether that specific PSU has been drawn previously or not. Hence, n independent draws are made.
Although sampling with replacement is less efficient than certain sampling schemes without replacement (see Brewer and Hanif (1983)), it is often preferred as

(i) selection of the PSUs is simple, since the drawing probabilities for each draw are the same;
(ii) any finite predetermined number, although not necessarily distinct, PSUs can be selected;
(iii) it produces a simple unbiased variance estimator; and
(iv) unbiased estimators of the total variance and components of variance in multistage sampling are easily obtained.

1.4.1 Single-stage cluster sampling

From the population of N PSUs, n PSUs are to be selected with replacement. If the i-th PSU is selected, all its elements are included in the sample.
On each of the n draws one PSU is selected with selection probability z_i, $i = 1, 2, \ldots, N$, with replacement. The selection probabilities are often taken as the relative measures of size, i.e. $z_i = M_i/M_0$. This will become apparent later.

For each draw an estimate of Y, the population total, can be obtained from each PSU selected. From (1.3) it can be seen that for any draw, Y can be estimated unbiasedly by

$$(\hat{Y})_i = \frac{y_i}{z_i} \tag{1.15}$$

if PSU i is selected.
Thus, after n draws, an unbiased estimator of the population total Y is given by

$$\hat{Y}_{pps} = \frac{1}{n}\sum_{i=1}^{n}(\hat{Y})_i \tag{1.16a}$$

$$= \frac{1}{n}\sum_{i=1}^{n}\frac{y_i}{z_i} \tag{1.16b}$$

According to Cochran (1977), the variance of the population total estimator is expressed as

$$Var(\hat{Y}_{pps}) = \frac{1}{n}\sum_{i=1}^{N} z_i\left(\frac{y_i}{z_i} - Y\right)^2 \tag{1.17a}$$

$$= \frac{1}{2n}\sum_{i=1}^{N}\sum_{j\neq i}^{N} z_i z_j\left(\frac{y_i}{z_i} - \frac{y_j}{z_j}\right)^2 \quad , \quad \text{or}$$

$$= \frac{1}{n}\sum_{i=1}^{N}\sum_{j>i}^{N} z_i z_j\left(\frac{y_i}{z_i} - \frac{y_j}{z_j}\right)^2 \tag{1.17b}$$

and an unbiased estimator of $Var(\hat{Y}_{pps})$ is given by

$$\hat{Var}(\hat{Y}_{pps}) = \frac{1}{n(n-1)}\sum_{i=1}^{n}\left(\frac{y_i}{z_i} - \hat{Y}_{pps}\right)^2 \tag{1.18a}$$

$$= \frac{1}{2n^2(n-1)}\sum_{i=1}^{n}\sum_{j\neq i}^{n}\left(\frac{y_i}{z_i} - \frac{y_j}{z_j}\right)^2 \quad , \quad \text{or}$$

$$= \frac{1}{n^2(n-1)}\sum_{i=1}^{n}\sum_{j>i}^{n}\left(\frac{y_i}{z_i} - \frac{y_j}{z_j}\right)^2 \quad . \tag{1.18b}$$

Consider the variance of $\hat{Y}_{pps}$ in (1.17a). It can be seen that this variance equals zero if $z_i = y_i/Y$. This means that the selection probabilities are directly proportional to the y_i's. However, the y_i's are unknown in practice and have to be observed. But, if it is known that some MOS, for each PSU i, is correlated with y_i, the selection probabilities can be chosen to be the relative MOS, i.e. $z_i = M_i/M_0$, $i = 1, 2, \ldots, N$. These selection probabilities will therefore also be correlated with the y_i's and the squared term in (1.17a) will be small, resulting in a reduced variance of $\hat{Y}_{pps}$. This result makes sampling with replacement, with selection probabilities relative to some MOS very appealing.

For the special case where the PSUs are selected with equal probabilities, i.e. simple random sampling, the selection probabilities are given by $z_i = 1/N$ for all i. This will also be the case if all the PSUs are of equal MOS and sampling with probabilities proportional to MOS is applied. Then, the estimator of the population total, (1.16b), becomes

$$\hat{Y}_{pps} = \frac{N}{n}\sum_{i=1}^{n} y_i \quad . \tag{1.19}$$

Its variance, (1.17b), becomes

$$Var(\hat{Y}_{pps}) = \frac{1}{n}\sum_{i=1}^{N}\sum_{j>i}^{N}(y_i - y_j)^2 \quad , \tag{1.20}$$

which will be estimated unbiasedly, using (1.18b), by

$$\hat{Var}(\hat{Y}_{pps}) = \frac{N^2}{n^2(n-1)} \sum_{i=1}^{n} \sum_{j>i}^{n} (y_i - y_j)^2 \quad . \tag{1.21}$$

1.4.2 Multistage cluster sampling

In multistage sampling, sampling of PSUs with replacement is often preferred as the selection and the estimation procedures are simple (see Brewer and Hanif (1983)).

The approach used in this section to estimate the variance of the population total is the one given in Brewer and Hanif (1983).

The variance of an unbiased estimator in multistage sampling is the sum of the variances arising from all the different stages, provided the selection and the estimation take place independently at the different stages. That is, the total variance of an unbiased estimator is the sum of the variance arising from the first sampling stage and the conditional variances arising from the second and later sampling stages.
Let $\hat{\theta}$ be an unbiased estimator of a population parameter θ. Then,

$$Var(\hat{\theta}) = Var_1(E_2(\hat{\theta})) + E_1(Var_2(\hat{\theta})) \quad , \tag{1.22}$$

where E_1 denotes the expected value and Var_1 the variance over all possible first stage samples, E_2 denotes the conditional expected value over all second and later stage samples, and Var_2 denotes the conditional variance of $\hat{\theta}$, given a particular first stage sample has been selected.

However, in sampling of PSUs with replacement, a given PSU can appear in the sample more than once. In this case an independent second stage sample of SSUs of the same size has to be selected each time a given PSU appears in the sample. This can be achieved in several ways. For example, say the i-th PSU has been selected twice, then the first subsample of SSUs that is to be selected from this PSU is replaced before the second independent subsample of SSUs is selected. Now, the component from the second and later stages of variance is exactly equal to the actual variance from these second and later stages. An unbiased estimator of the total variance is obtained by replacing the total of each PSU by an unbiased estimator of these totals.

As in the case of single-stage sampling, suppose that n of the N PSUs are to be selected with replacement with probability of selecting PSU i on any draw being z_i, $i = 1, 2, \ldots, N$. Suppose that the subsampling in each of the n selected PSUs satisfy the above mentioned requirements. That is, at each of the second and later sampling stages the subsamples are selected independently and an independent, unbiased estimator of the population total is obtained for each subsample. Whenever a certain PSU is selected more than once, an independent subsample of SSUs of the same size is selected each time that PSU appears in the sample.

Let Y_i denote the PSU-total for the i-th PSU, which is the total of y over all subunits contained in PSU i. Now, for each of the n PSUs selected, the PSU-total, Y_i, is estimated unbiasedly by the second and later stage estimator $\hat{Y}_i$.

Thus, as in (1.3), for each of the n draws, Y, the population total, is estimated unbiasedly by

$$(\hat{Y})_i = \frac{\hat{Y}_i}{z_i} \tag{1.23}$$

if the i-th PSU is selected. Note that the n estimates $(\hat{Y})_i$ are independent.
After n draws, an unbiased estimator of the population total Y is given by

$$\begin{aligned} \hat{Y}_{pps} &= \frac{1}{n}\sum_{i=1}^{n}(\hat{Y})_i \\ &= \frac{1}{n}\sum_{i=1}^{n}\frac{\hat{Y}_i}{z_i} \end{aligned} \quad , \tag{1.24}$$

which can be seen to be equivalent to (1.16b) in the case of single-stage sampling.

The variance of $\hat{Y}_{pps}$ is given by

$$Var(\hat{Y}_{pps}) = \frac{1}{n}\sum_{i=1}^{N} z_i \left(\frac{Y_i}{z_i} - Y\right)^2 + \frac{1}{n}\sum_{i=1}^{N}\frac{Var_2(\hat{Y}_i)}{z_i} \tag{1.25a}$$

$$= \frac{1}{n}\sum_{i=1}^{N}\sum_{j>i}^{N} z_i z_j \left(\frac{Y_i}{z_i} - \frac{Y_j}{z_j}\right)^2 + \frac{1}{n}\sum_{i=1}^{N}\frac{Var_2(\hat{Y}_i)}{z_i} \quad , \tag{1.25b}$$

where $Var_2(\hat{Y}_i)$ is the second and later stage variance of $\hat{Y}_i$.

Note that the first term in (1.25a) is the variance of $\hat{Y}_{pps}$ when each PSU is completely enumerated and is the same as (1.17a) for single-stage sampling if Y_i is replaced by y_i in (1.25a). It can thus be seen that (1.25a) takes on the form of (1.22).

Recall the variance estimator of $Var(\hat{Y}_{pps})$ in single-stage sampling as given by (1.18a).Now the variance estimator formed by replacing y_i by $\hat{Y}_i$ in (1.18a) has expectation

$$
\begin{aligned}
&E\left[\frac{1}{n(n-1)}\sum_{i=1}^{n}\left(\frac{\hat{Y}_i}{z_i}-\hat{Y}_{pps}\right)^2\right] \\
&=\frac{1}{n}\sum_{i=1}^{N} z_i\left(\frac{Y_i}{z_i}-Y\right)^2+\frac{1}{n}\sum_{i=1}^{N}\frac{Var_2(\hat{Y}_i)}{z_i} \\
&=Var(\hat{Y}_{pps}) \qquad .
\end{aligned}
$$

Thus,

$$\hat{Var}(\hat{Y}_{pps}) = \frac{1}{n(n-1)}\sum_{i=1}^{n}\left(\frac{\hat{Y}_i}{z_i}-\hat{Y}_{pps}\right)^2 \tag{1.26a}$$

$$= \frac{1}{n^2(n-1)}\sum_{i=1}^{n}\sum_{j>i}^{n}\left(\frac{\hat{Y}_i}{z_i}-\frac{\hat{Y}_j}{z_j}\right)^2 \tag{1.26b}$$

is an unbiased estimator of $Var(\hat{Y}_{pps})$, the total variance of $\hat{Y}_{pps}$.

This implies that the variance of the population total Y can be estimated by taking only the first sampling stage into consideration.

Consider the variance of $\hat{Y}_{pps}$ in (1.25a). It can be seen that the first term in (1.25a) equals zero if the selection probabilities are directly proportional to the Y_i's, that is, if $z_i = Y_i/Y$ for all i. This means that there is no first-stage variance and the total variance is equal to the variances arising from the second and later stages. Thus the total variance is reduced. However, as in the case of single-stage sampling, the Y_i's are unknown in practice. If it is known that some measure of size is correlated with Y_i, the selection probabilities defined by $z_i = M_i/M_0$ will also be correlated with Y_i. The first term in (1.25a) will thus be small, resulting in a reduced variance in the first sampling stage. Therefore, the total variance of $\hat{Y}_{pps}$ will be reduced.

It has thus been shown that sampling of PSUs with replacement produces a simple unbiased variance estimator of the total variance. However this is only the case if the sampling schemes used in the second and later stages meet the above mentioned requirements. If any other sampling schemes are applied, the estimator of the total variance ceases to be unbiased. It is thus essential that $\hat{Y}_i$ is an unbiased estimator of the i-th PSU-total, Y_i, such that $(\hat{Y})_i$ are n unbiased, independent estimators of the population total Y.

Similar to single-stage sampling, if sampling of PSUs takes place with equal probabilities, i.e. simple random sampling, (1.24) is now given by

$$\hat{Y}_{pps} = \frac{N}{n}\sum_{i=1}^{n}\hat{Y}_i \quad . \tag{1.27}$$

By using (1.25b), its variance is given by

$$Var(\hat{Y}_{pps}) = \frac{1}{n}\sum_{i=1}^{N}\sum_{j>i}^{N}(Y_i - Y_j)^2 + \frac{N}{n}\sum_{i=1}^{N}Var_2(\hat{Y}_i) \quad . \tag{1.28}$$

By using (1.26b), an unbiased estimator of $Var(\hat{Y}_{pps})$ is given by

$$\hat{Var}(\hat{Y}_{pps}) = \frac{N^2}{n^2(n-1)}\sum_{i=1}^{n}\sum_{j>i}^{n}(\hat{Y}_i - \hat{Y}_j)^2 \quad . \tag{1.29}$$

1.5 SAMPLING OF PSUs WITHOUT REPLACEMENT

In sampling without replacement, for a sample of size n, one PSU is selected at each of the n draws, but once selected, a specific PSU cannot be drawn again. The n draws are thus not independent, but a sample of n distinct PSUs is obtained, which is not necessarily the case for sampling with replacement.

1.5.1 Single-stage cluster sampling

Consider N PSUs of MOS M_i, where $M_0 = \sum_{i=1}^{N} M_i$ is the total MOS of the N PSUs in the population. A sample of n PSUs is to be selected by some method without replacement from the N PSUs in the population. If the i-th PSU is selected, all its elements are included in the sample.

The initial selection probabilities of the N PSUs are defined by z_i, $i = 1, 2, \ldots, N$. Often, these selection probabilities are taken as $z_i = M_i/M_0$. This will become apparent later. Suppose that the inclusion probabilities, π_i, are known.

By using (1.3), it can be seen that for any draw the population total Y can be estimated by simply dividing y_i, the item total of the variable under investigation of the i-th PSU selected, by the probability of selecting PSU i on that draw. For n draws, the simple mean of the n estimates can be used to estimate the population total. This can be seen from (1.16a) in the case of sampling with replacement .

Now, for sampling without replacement, $\overline{\pi}_i$, the average probability, is the probability attached to PSU i of being drawn on each of the n draws.

Then, by using (1.16a), Y can be estimated by

$$\begin{aligned}\hat{Y}_{HT} &= \frac{1}{n}\left(\frac{y_1}{\overline{\pi}_1} + \frac{y_2}{\overline{\pi}_2} + \ldots + \frac{y_n}{\overline{\pi}_n}\right) \\ &= \frac{1}{n}\sum_{i=1}^{n}\left(\frac{y_i}{\overline{\pi}_i}\right) \quad ,\end{aligned} \tag{1.30a}$$

or, since $\pi_i = n\overline{\pi}_i$,

$$\hat{Y}_{HT} = \sum_{i=1}^{n} \frac{y_i}{\pi_i} \quad . \tag{1.30b}$$

This estimator is unbiased if

$$\overline{\pi}_i > 0 \quad , \quad 1 \leq i \leq N \quad \text{and} \quad \sum_{i=1}^{N} \overline{\pi}_i = 1 \quad ,$$

which implies that

$$\sum_{i=1}^{N} \pi_i = n \quad .$$

Since this estimator was first proposed by Horvitz and Thompson (1952), it is called the Horvitz-Thompson estimator.

Brewer and Hanif (1983) give the following important properties of the Horvitz-Thompson estimator:

(i) It is the only unbiased estimator of all the estimators of the population total that attach the same weight to a particular population element whenever it is selected.
(ii) $\hat{Y}_{HT}$ has the smallest variance of all the population total estimators which are homogeneous linear unbiased.
(iii) The variance of $\hat{Y}_{HT}$ is zero, provided the π_i's are exactly proportional to the y_i's and the sample size is fixed.

Property (iii) is very important in the choice of sampling scheme applied. Suppose that the initial selection probability, z_i, is directly proportional to y_i, for all i. Thus, if the inclusion probabilities are given by $\pi_i = nz_i$, they are also directly proportional to y_i, for all i. Then it can be seen from (1.30b) that

$$\hat{Y}_{HT} = \sum_{i=1}^{n} \frac{y_i}{\pi_i} = \frac{1}{n} \sum_{i=1}^{n} \frac{y_i}{z_i}$$

will have zero variance, since then $z_i = y_i/Y$ and for each PSU selected, Y will be estimated exactly, namely by

$$\frac{1}{n} \sum_{i=1}^{n} \frac{y_i}{z_i} = \frac{1}{n} \sum_{i=1}^{n} Y = Y \quad .$$

According to Jessen (1978), regardless of the selection scheme applied, as long as

$$0 < \pi_i \leq 1 \quad \text{and} \quad \sum_{i=1}^{N} \pi_i = n \quad ,$$

the variance of $\hat{Y}_{HT}$ is given by

$$Var(\hat{Y}_{HT}) = \sum_{i=1}^{N} \frac{(1-\pi_i)}{\pi_i} y_i^2 + 2\sum_{i=1}^{N}\sum_{j>i}^{N} \frac{(\pi_{ij}-\pi_i\pi_j)}{\pi_i\pi_j} y_i y_j \quad . \tag{1.31}$$

An unbiased estimator of $Var(\hat{Y}_{HT})$ (see Cochran (1977)) is then given by

$$\hat{Var}(\hat{Y}_{HT}) = \sum_{i=1}^{n} \frac{(1-\pi_i)}{\pi_i^2} y_i^2 + 2\sum_{i=1}^{n}\sum_{j>i}^{n} \frac{(\pi_{ij}-\pi_i\pi_j)}{\pi_i\pi_j\pi_{ij}} y_i y_j \quad , \tag{1.32}$$

provided $\pi_{ij} \neq 0$.
However, this estimator is not always zero when the variance is zero.

Yates and Grundy (1953) and according to Brewer and Hanif (1983), also Sen (1953) derived an alternative expression for the variance of $\hat{Y}_{HT}$, namely,

$$Var(\hat{Y}_{HT}) = \sum_{i=1}^{N}\sum_{j>i}^{N} (\pi_i\pi_j - \pi_{ij})(\frac{y_i}{\pi_i} - \frac{y_j}{\pi_j})^2 \quad , \tag{1.33}$$

which is only valid if the sample size is fixed, which means that relations (1.13) hold. To express $\hat{Var}(\hat{Y}_{HT})$ in the more compact form of (1.33), the first two of these relations are used.
According to Brewer and Hanif (1983), a conditional unbiased estimator, i.e. if $\pi_{ij} \neq 0$, of $Var(\hat{Y}_{HT})$ as given by Yates and Grundy (1953) and Sen (1953) for a fixed sample size is defined as

$$\hat{Var}(\hat{Y}_{HT}) = \sum_{i=1}^{n}\sum_{j>i}^{n} \frac{(\pi_i\pi_j - \pi_{ij})}{\pi_{ij}} (\frac{y_i}{\pi_i} - \frac{y_j}{\pi_j})^2 \quad . \tag{1.34}$$

It has already been shown that if the sampling scheme is chosen such that the inclusion probabilities are given by $\pi_i = nz_i$ for all i, and the initial selection probabilities are given by $z_i = y_i/Y$, that $Var(\hat{Y}_{HT})$ will be zero. However the y_i's are unknown in practice. But if it is known that some MOS is correlated with y_i and the initial selection probabilities are chosen to be the relative measures of size, i.e. $z_i = M_i/M_0$ for all i, these selection probabilities will also be correlated with y_i. Thus, the inclusion probabilities given by $\pi_i = nz_i$ will also be correlated with y_i for all i, and the variance (1.33) will be reduced.

Both variance estimators, (1.32) and (1.34), may take on negative values, although (1.34) does so rarely (Brewer and Hanif (1983)). Vijayan (1975) has shown that (1.34) is the only possible nonnegative estimator for a sample of size two. Durbin (1967) states that the factor $(\pi_i\pi_j - \pi_{ij})/\pi_{ij}$ in (1.34) varies from sample to sample and may take on large values, which causes an instability in the estimate of the variance. From inspection of (1.34) it is obvious that $\hat{Var}(\hat{Y}_{HT})$ will be positive and more stable if the sampling scheme adopted assures that $\pi_i\pi_j - \pi_{ij} > 0$, for all i and j, $i \neq j$, (Sunter(1986)). Jessen (1978) states that the variance estimator (1.34) will be positive if the sampling scheme applied, makes π_{ij} proportionate to $\pi_i\pi_j$.

It is thus clear that sampling without replacement is very sensitive to the sampling scheme applied.

For the special case where the PSUs are selected with equal probabilities, i.e. simple random sampling, the selection probabilities are given by $z_i = 1/N$ for all i. This will also be the case if all the PSUs are of equal MOS and sampling with probabilities proportional to MOS is applied. It is well known that the inclusion probabilities and joint inclusion probabilities are then given by

$$\pi_i = \frac{n}{N} = nz_i$$

$$\text{and} \qquad \pi_{ij} = \frac{n(n-1)}{N(N-1)} \quad , \quad i \neq j \quad ,$$

respectively.

Then, the estimator of the population total, (1.30b), becomes

$$\hat{Y}_{HT} = \frac{N}{n}\sum_{i=1}^{n} y_i \quad . \tag{1.35}$$

Its variance, (1.33), becomes

$$Var(\hat{Y}_{HT}) = \frac{N-n}{n(N-1)}\sum_{i=1}^{N}\sum_{j>i}^{N}(y_i - y_j)^2 \quad , \tag{1.36}$$

which can be estimated unbiasedly, using (1.34), by

$$\hat{Var}(\hat{Y}_{HT}) = \frac{N(N-n)}{n^2(n-1)}\sum_{i=1}^{n}\sum_{j>i}^{n}(y_i - y_j)^2 \quad . \tag{1.37}$$

1.5.2 Multistage cluster sampling

According to Brewer and Hanif (1983) and also Cochran (1977), Rao (1975) developed the following analysis to derive general formulae for the estimation of the total variance for multistage sampling designs, which is especially useful in without replacement designs:

Suppose that n of the N PSUs are selected by some or other method without replacement and that subsampling in each of the n selected PSUs takes place independently. Denote the sample of n PSUs by s. Thus for each selected PSU an independent, unbiased estimator, $\hat{Y}_i$, based on subsampling in PSU i, of the i-th PSU-total is obtained.

The population total, Y, can be estimated unbiasedly by a linear multistage estimator of the form

$$\hat{Y} = \sum_{i \in s} a_{is} \hat{Y}_i \quad , \tag{1.38}$$

where a_{is} is a weight known for each sample s and may depend on the sample s as well as on PSU i.

Let a'_{is} be a random variable defined by

$$a'_{is} = \begin{cases} a_{is} & \text{if} \quad s \ni i \\ 0 & \text{otherwise.} \end{cases}$$

Thus,

$$\hat{Y} = \sum_{i=1}^{N} a'_{is} \hat{Y}_i \quad . \tag{1.39}$$

Now,

$$E(\hat{Y}) = E_1 E_2(\hat{Y}) = E_1 \left(\sum_{i=1}^{N} a'_{is} Y_i \right) = Y \quad ,$$

if and only if $E_1(a'_{is}) = 1$ for all i.

Define

$$\hat{Y}_c = \sum_{i=1}^{N} a'_{is} Y_i$$

as the unbiased estimator of the population total, Y, for single-stage sampling.

The variance of $\hat{Y}$ is then given by

$$Var(\hat{Y}) = Var(\hat{Y}_c) + E_1 \left[\sum_{i \in s} a^2_{is} Var_{2s}(\hat{Y}_i) \right] \quad , \tag{1.40}$$

where $Var_{2s}(\hat{Y}_i)$ is the conditional variance of $\hat{Y}_i$, given a certain sample s of PSUs.

Suppose that an unbiased estimator for $Var(\hat{Y}_c)$ exists. This estimator will be of the form

$$\hat{Var}(\hat{Y}_c) = \sum_{i \in s} b_{is} Y_i^2 + \sum_{i<j} \sum_{\in s} d_{is} Y_i Y_j \quad ,$$

where b_{is} and d_{is} are weights known for each sample s.

Also, suppose that there exists an unbiased estimator for $Var_2(\hat{Y}_i)$, namely $\hat{Var}_{2s}(\hat{Y}_i)$. If $Var_{2s}(\hat{Y}_i)$ is independent of s, it can be written as $Var_2(\hat{Y}_i)$. According to Brewer and Hanif (1983), Raj (1966) proved that for this case, an unbiased estimator of $Var(\hat{Y})$ is given by

$$\hat{Var}_A(\hat{Y}) = \sum_{i\in s} b_{is}\hat{Y}_i^2 + \sum_{i,j}\sum_{\in s} d_{is}\hat{Y}_i\hat{Y}_j + \sum_{i\in s} a_{is}\hat{Var}_2(\hat{Y}_i) \quad . \tag{1.41}$$

Regardless whether $Var_{2s}(\hat{Y}_i)$ is independent of s or not, an alternative unbiased estimator of $Var(\hat{Y})$ was given by Rao (1975), namely,

$$\hat{Var}_B(\hat{Y}) = \sum_{i\in s} b_{is}\hat{Y}_i^2 + \sum_{i<j}\sum_{\in s} d_{is}\hat{Y}_i\hat{Y}_j + \sum_{i\in s}(a_{is}^2 - b_{is})\hat{Var}_{2s}(\hat{Y}_i) \quad . \tag{1.42}$$

The multistage analogue of the Horvitz-Thompson estimator is defined by

$$\hat{Y}_{HT} = \sum_{i=1}^{n} \frac{\hat{Y}_i}{\pi_i} \tag{1.43}$$

where the weight a'_{is} is given by

$$a'_{is} = \begin{cases} \frac{1}{\pi_i} & \text{if} \quad s \ni i \\ 0 & \text{otherwise} \end{cases}$$

and where π_i is the inclusion probability of the i-th PSU in the sample and $\sum_{i=1}^{n}\pi_i = n$. $\hat{Y}_{HT}$ is thus an unbiased estimator of the population total Y.

According to Durbin (1953), the variance of the Horvitz-Thompson estimator is given by

$$Var(\hat{Y}_{HT}) = \sum_{i=1}^{N}\pi_i(1-\pi_i)\left(\frac{Y_i}{\pi_i}\right)^2 - \sum_{i=1}^{N}\sum_{j\neq i}^{N}(\pi_i\pi_j - \pi_{ij})\left(\frac{Y_i}{\pi_i}\right)\left(\frac{Y_j}{\pi_j}\right) + \sum_{i=1}^{N}\frac{Var_{2s}(\hat{Y}_i)}{\pi_i} \quad , \tag{1.44a}$$

which can also be written in the form

$$Var(\hat{Y}_{HT}) = \sum_{i=1}^{N}\sum_{j>i}^{N}(\pi_i\pi_j - \pi_{ij})\left(\frac{Y_i}{\pi_i} - \frac{Y_j}{\pi_j}\right)^2 + \sum_{i=1}^{N}\frac{Var_{2s}(\hat{Y}_i)}{\pi_i} \quad . \tag{1.44b}$$

According to Brewer and Hanif (1983), Brewer and Hanif (1970) derived the Sen-Yates-Grundy Variance Estimator for the Horvitz-Thompson Estimator from (1.42), namely,

$$\hat{Var}(\hat{Y}_{HT}) = \sum_{i=1}^{n}\sum_{j>i}^{n}\frac{\pi_i\pi_j - \pi_{ij}}{\pi_{ij}}\left(\frac{\hat{Y}_i}{\pi_i} - \frac{\hat{Y}_j}{\pi_j}\right)^2 + \left[\frac{1}{n-1} - \frac{\pi_i\pi_j - \pi_{ij}}{\pi_{ij}}\right]\left[\frac{\hat{Var}_{2s}(\hat{Y}_i)}{\pi_i^2} + \frac{\hat{Var}_{2s}(\hat{Y}_j)}{\pi_j^2}\right] \quad . \tag{1.45}$$

From (1.44b) it is clear that if $\pi_i = nY_i/Y$, i.e. π_i is directly proportional to Y_i, that the first term in (1.44b) will be zero and the first-stage variance vanishes. The total variance will thus be reduced, as it now only equals the variance arising from the second and later stages.

However, the Y_i's are unknown. If some MOS is correlated with Y_i, the inclusion probabilities given by $\pi_i = nz_i$, where $z_i = M_i/M_0$, are also correlated with the Y_i's. Then the squared term in (1.44b) will be reduced, resulting in a reduced variance of $\hat{Y}_{HT}$.

Consider the variance estimator of $\hat{Y}_{HT}$ as given by (1.45). In contrast to sampling with replacement, where the variance of the population total can be estimated by only taking the first sampling stage into consideration (see (1.26a) and (1.26b)), here the variances arising from the second and later stages have to be taken into consideration in order to estimate the total variance of $\hat{Y}_{HT}$.

For example, let n PSUs be selected without replacement, where π_i is the inclusion probability of PSU i. Suppose that the i-th PSU contains M_i elements of which m_i are selected by simple random sampling without replacement, when PSU i is drawn.

According to Cochran (1977), the variance of $\hat{Y}_{HT}$ is then estimated by

$$\hat{Var}(\hat{Y}_{HT}) = \sum_{i=1}^{n}\sum_{j>i}^{n} \frac{\pi_i\pi_j - \pi_{ij}}{\pi_{ij}} \left(\frac{\hat{Y}_i}{\pi_i} - \frac{\hat{Y}_j}{\pi_j}\right)^2 + \sum_{i=1}^{n} \frac{M_i^2(1-f_{2i})s_{2i}^2}{m_i\pi_i} \quad ,$$

where

$$f_{2i} = \frac{m_i}{M_i} \qquad \text{and} \qquad s_{2i}^2 = \sum_{j=1}^{m_i} \frac{(y_{ij} - \bar{y}_i)^2}{m_i - 1} \quad ,$$

where y_{ij} is the observation of the j-th subunit within the i-th PSU and $\bar{y}_i$ is the mean of y per subunit.

If a third sampling stage is added, another term estimating the variance introduced by this stage has to be added in order to estimate the total variance of $Var(\hat{Y}_{HT})$.

Similarly to single-stage sampling, if sampling of PSUs takes place with equal probabilities, i.e. simple random sampling, (1.43) is now given by

$$\hat{Y}_{HT} = \frac{N}{n}\sum_{i=1}^{n} \hat{Y}_i \tag{1.46}$$

and, by using (1.44b), its variance is given by

$$Var(\hat{Y}_{HT}) = \frac{N-n}{n(N-1)}\sum_{i=1}^{N}\sum_{j>i}^{N}(Y_i - Y_j)^2 + \frac{N}{n}\sum_{i=1}^{N} Var_{2s}(\hat{Y}_i) \quad , \tag{1.47}$$

which is estimated unbiasedly, using (1.45), by

$$\hat{Var}(\hat{Y}_{HT}) = \frac{N(N-n)}{n^2(n-1)}\sum_{i=1}^{n}\sum_{j>i}^{n}(\hat{Y}_i - \hat{Y}_j)^2 + \frac{N}{n(n-1)}\left[\hat{Var}_{2s}(\hat{Y}_i) + \hat{Var}_{2s}(\hat{Y}_j)\right] \quad . \tag{1.48}$$

1.6 THE PROBLEM

The problem concerns the estimation of the variance of the population total estimator.

Firstly consider single-stage sampling. By comparing the variance estimators, (1.18b) for sampling with replacement and (1.34) for sampling without replacement, it is clear that the first is more convenient than the latter, since the first is free of the inclusion probabilities, π_i, and the joint inclusion probabilities, π_{ij}. The π_i's and π_{ij}'s which have to be known for sampling without replacement, are quite troublesome since:

(i) π_i and π_{ij} are often difficult to calculate, depending on the sampling scheme used. The sampling procedures should permit calculating π_{ij} readily, so that it is not necessary to determine all possible combinations of samples that contain both units i and j, as was done in the example in section 1.3.

(ii) The sampling scheme adopted should ensure that $\pi_{ij} \neq 0$ for all $i, j (i \neq j)$ in order to be able to estimate the variance estimator.

(iii) For the variance estimator to be positive and more stable, it is essential that $\pi_i \pi_j - \pi_{ij} > 0$ for all $i, j (i \neq j)$. For a sample of size 2, irrespective of the sampling scheme, this relation does hold.

(iv) Hajek (1981) states that for many interesting sampling schemes, the relation $\pi_i \pi_j - \pi_{ij}$ is difficult to find exactly.

Now consider multistage sampling. By comparing the variance estimators, (1.26b) for sampling with replacement and (1.45) for sampling without replacement, it can be seen that the first is more convenient than the latter, since for the latter:

(i) as in the case of single-stage sampling, it involves the troublesome inclusion probabilities, π_i, and joint inclusion probabilities, π_{ij} ; and

(ii) it requires the separate calculation of the variance due to sampling at the second and later stages, which is not the case for sampling with replacement.

This second property is highly undesirable, since the sample designs at the various stages have to be known and for each stage the variance has to be estimated according to the sampling design applied. This is rather a tedious task, especially if many sampling stages are used. Also, if at any sampling stage further stratification is necessary, this task will be even more complicated.

Thus for sampling of PSUs without replacement, for each different complex sample design, a different variance estimator depending on the sampling designs in the second and later stages, has to be used in order to estimate the total variance of the population total

estimator. For sampling of PSUs with replacement one unique estimator, independent of the sampling designs in the second and later stages, can be used.
If the variance is estimated with the aid of a computer, for each complex sample design, a new program has to be designed for sampling without replacement, whereas for sampling with replacement the same program can be used.

Thus, whether single-stage or multistage sampling is used, it is clear that sampling of PSUs with replacement is more convenient than sampling of PSUs without replacement.

Sampling with replacement however is often undesirable as it permits a PSU to be included in the sample more than once. In complex sampling this can easily happen as the number of primary clusters per stratum is usually not large. It can even be less than 10. Durbin (1967) states: "The fact that repetitions are possible in with-replacement sampling is, I believe, indicative of the inefficiency of the method."
To prevent duplications in the sample, the sample can be selected either without replacement or, otherwise, by using a rejective method. A rejective method is a sampling scheme where the PSUs are selected with replacement, but as soon as a duplication appears in the sample, the partially built up sample is rejected and a new sample is then drawn. This process is continued until a sample of n distinct PSUs is obtained. Konijn (1973) states that for a rejective sampling scheme, where a fixed number of distinct PSUs in the sample is assured, the distinction between sampling of PSUs with and without replacement loses practical relevance. Since relations (1.13) also hold for rejective sampling schemes, which will be shown by means of examples later, the population total estimator, its variance and variance estimator are identical to those used for sampling without replacement. See also Yates and Grundy (1953) and Hajek (1981) in this respect. Thus results obtained for sampling without replacement also hold for rejective sampling methods.

In order to overcome the above mentioned problems encountered with the estimation of the variance when sampling of PSUs takes place without replacement, but no duplications in the sample are desired, it is often customary to select the primary units without replacement or by using a rejective method, but for an estimate of the variance, the formula based on sampling with replacement is used instead of the formula based on sampling without replacement.
This means, sampling without replacement is applied, but the population total is estimated by $\hat{Y}_{pps}$ and the variance of $\hat{Y}_{pps}$ is estimated by $\hat{Var}_{wr}(\hat{Y}_{pps})$, where the suffix wr denotes sampling with replacement.

Kish (1965) states in this regard: " The primary units may be selected with replacement, to

suit the variance formulas. But often we may select them without replacement and accept a mild overestimate of the variance. That is, by selecting primary units with replacement, the actual variance will probably be made a little lower than the formulas show."
This assertion means that $\hat{Var}_{wr}(\hat{Y}_{pps})$ probably overestimates the true variance of $\hat{Y}_{pps}$ under sampling without replacement. Kish gives no substantiation for his assertion and the validity of the assertion is not self-evident. The aim of this study is to investigate the general validity of the Kish assertion or, otherwise, the requirements under which the assertion is true.

If N, the number of PSUs in a stratum is large and n, the sample size is small, sampling without replacement can in effect be approximated by sampling with replacement so that the Kish assertion is at least approximately true. However, in complex sampling schemes the sampling fraction n/N is usually not negligible. Samples of size two are often selected to ensure maximum stratification, but a stratum usually contains 10 to 15 PSUs and N may even be as small as 8.

If for sampling without replacement, the sampling scheme applied ensures that $\pi_i = nz_i$, then

$$\hat{Y}_{HT} = \sum_{i=1}^{n} \frac{y_i}{\pi_i}$$
$$= \frac{1}{n} \sum_{i=1}^{n} \frac{y_i}{z_i} = \hat{Y}_{pps} \quad ,$$

i.e. the two population estimators are identical and $\hat{Y}_{pps}$ is also an unbiased estimator of the population total under sampling without replacement. This is also true for multistage sampling. Since $\hat{Y}_{pps}$ is unbiased, the true variance of $\hat{Y}_{pps}$ is given by $Var_{wor}(\hat{Y}_{pps})$, where suffix wor denotes sampling without replacement. For this case, Rao (1988) states that $\hat{Var}_{wr}(\hat{Y}_{pps})$ overestimates $Var_{wor}(\hat{Y}_{pps})$. However, this is also not self-evident and the effect of using the inappropriate variance estimator will be investigated in Chapter 2.

If for sampling without replacement, $\pi_i \neq nz_i$, the two population total estimators are not identical. Thus the use of $\hat{Y}_{pps}$ will lead to a biased estimator of the population total. In this case, the mean square error instead of the variance of $\hat{Y}_{pps}$ should be estimated. Denote the mean square error of $\hat{Y}_{pps}$ under sampling without replacement by $MSE_{wor}(\hat{Y}_{pps})$. It is not self-evident that $\hat{Var}_{wr}(\hat{Y}_{pps})$ overestimates $MSE_{wor}(\hat{Y}_{pps})$. The effect of this method will be investigated in Chapter 3.

In Chapter 4 various sampling procedures for which a sample of n distinct PSUs is obtained will be compared. A recommendation will be made of which procedure should be applied in practice.

CHAPTER 2

SAMPLING WITHOUT REPLACEMENT IN THE CASE WHERE $\pi_i = nz_i$

2.1 THE EFFECT OF AN INAPPROPRIATE VARIANCE ESTIMATOR

2.1.1 Introduction

As has been shown in Section 1.6, if the sampling scheme applied for sampling without replacement ensures that the inclusion probabilities are given by $\pi_i = nz_i$, $\hat{Y}_{HT}$ is identical to $\hat{Y}_{pps}$ and the population total is estimated unbiasedly. In this case the true variance of $\hat{Y}_{pps}$ under sampling without replacement is given by $Var_{wor}(\hat{Y}_{pps})$. An unbiased estimator of this variance would be $\hat{Var}_{wor}(\hat{Y}_{pps})$, but Kish recommended that $\hat{Var}_{wr}(\hat{Y}_{pps})$ should be used. However in order to ensure that $\pi_i = nz_i$ in sampling without replacement , the drawing probabilities for each draw do not necessarily equal z_i any more. Therefore $\hat{Var}_{wr}(\hat{Y}_{pps})$ is a biased estimator of the variance $Var_{wor}(\hat{Y}_{pps})$.

Now in order for the Kish assertion to be true, it seems reasonable that the relation

$$E_{wor}\left[\hat{Var}_{wr}(\hat{Y}_{pps})\right] > Var_{wor}(\hat{Y}_{pps}) \tag{2.1}$$

should hold. This means that the use of $\hat{Var}_{wr}(\hat{Y}_{pps})$ will result on the average in a conservative estimate.

In this paragraph, relation (2.1) will be investigated separately for single-stage and multi-stage sampling.

2.1.2 Single-stage sampling

Recall again the variances of the population total and its estimators for sampling with and without replacement as given in Sections 1.4.1 and 1.5.1 respectively:

$$Var_{wr}(\hat{Y}_{pps}) = \frac{1}{n}\sum_{i=1}^{N}\sum_{j>i}^{N} z_i z_j \left(\frac{y_i}{z_i} - \frac{y_j}{z_j}\right)^2 \tag{2.2}$$

and

$$\hat{Var}_{wr}(\hat{Y}_{pps}) = \frac{1}{n^2(n-1)}\sum_{i=1}^{n}\sum_{j>i}^{n}\left(\frac{y_i}{z_i} - \frac{y_j}{z_j}\right)^2 \tag{2.3}$$

for sampling with replacement and

$$Var_{wor}(\hat{Y}_{pps}) = \sum_{i=1}^{N}\sum_{j>i}^{N}(\pi_i\pi_j - \pi_{ij})(\frac{y_i}{\pi_i} - \frac{y_j}{\pi_j})^2 \qquad (2.4)$$

and $$\hat{Var}_{wor}(\hat{Y}_{pps}) = \sum_{i=1}^{n}\sum_{j>i}^{n}\frac{\pi_i\pi_j - \pi_{ij}}{\pi_{ij}}(\frac{y_i}{\pi_i} - \frac{y_j}{\pi_j})^2 \qquad (2.5)$$

for sampling without replacement .

If the selection probabilities, z_i, for sampling with replacement are such that they equal $1/n$ times the inclusion probabilities, π_i, for sampling without replacement , (2.2) and (2.3) become

$$Var_{wr}(\hat{Y}_{pps}) = \frac{1}{n}\sum_{i=1}^{N}\sum_{j>i}^{N}\pi_i\pi_j(\frac{y_i}{\pi_i} - \frac{y_j}{\pi_j})^2 \qquad (2.6)$$

and

$$\hat{Var}_{wr}(\hat{Y}_{pps}) = \frac{1}{n-1}\sum_{i=1}^{n}\sum_{j>i}^{n}(\frac{y_i}{\pi_i} - \frac{y_j}{\pi_j})^2 \quad , \qquad (2.7)$$

respectively.

Durbin (1953) has shown that $\hat{Var}_{wr}(\hat{Y}_{pps})$ as an estimate of $Var_{wor}(\hat{Y}_{pps})$ has a bias of $n/(n-1)$ times the reduction in variance obtained by sampling without replacement instead of with replacement, i.e.

$$\text{bias} = \frac{n}{n-1}\left[Var_{wr}(\hat{Y}_{pps}) - Var_{wor}(\hat{Y}_{pps})\right] \quad .$$

Therefore

$$E_{wor}\left[\hat{Var}_{wr}(\hat{Y}_{pps})\right] = Var_{wor}(\hat{Y}_{pps}) + \frac{n}{n-1}\left[Var_{wr}(\hat{Y}_{pps}) - Var_{wor}(\hat{Y}_{pps})\right] \quad . \qquad (2.8)$$

Thus if the difference

$$Var_{wr}(\hat{Y}_{pps}) - Var_{wor}(\hat{Y}_{pps}) \qquad (2.9)$$

is positive, relation (2.1) holds and the variance is overestimated. From the factor $n/(n-1)$ in (2.8), it can be seen that the smaller the n, the larger the bias will be if (2.9) stays constant. Especially, if $n = 2$, as is often the case in complex sampling, this bias will be twice the reduction in variance.

It was shown in Chapter 1 that if the selection probabilities are directly proportional to the y_i's, $Var_{wr}(\hat{Y}_{pps}) = 0$ for sampling with replacement and $Var_{wor}(\hat{Y}_{pps}) = 0$ for sampling without replacement , when $\pi_i = nz_i$.

It was also shown in Chapter 1 that if a measure of size is correlated with y_i and the initial selection probabilities are taken as the relative measures of size, i.e. $z_i = M_i/M_0$, $Var_{wr}(\hat{Y}_{pps})$ and also $Var_{wor}(\hat{Y}_{pps})$, where $\pi_i = nz_i$, are small. Thus the difference (2.9) is small and the bias will be reduced. Thus, if it is known that $Var_{wr}(\hat{Y}_{pps})$ exceeds $Var_{wor}(\hat{Y}_{pps})$, the inappropriate variance estimator $\hat{Var}_{wr}(\hat{Y}_{pps})$ under sampling without replacement will on the average overestimate the true variance $Var_{wor}(\hat{Y}_{pps})$ to a lesser degree.

Sampling with selection probabilities proportional to some measure of size that is expected to be correlated with y is thus advisable.

However it is not possible to show that $Var_{wr}(\hat{Y}_{pps})$ exceeds $Var_{wor}(\hat{Y}_{pps})$ in the general case where $\pi_i = nz_i$ for sampling without replacement , since $Var_{wor}(\hat{Y}_{pps})$ also depends on the joint inclusion probabilities, π_{ij}, and the values of these quantities depend on the particular sampling scheme applied in sampling without replacement .

By comparing (2.6) for sampling with replacement and (2.4) for sampling without replacement , it is clear that

$$Var_{wr}(\hat{Y}_{pps}) > Var_{wor}(\hat{Y}_{pps})$$

$$\text{if} \quad \pi_{ij} > \pi_i \pi_j \frac{n-1}{n} \quad , \quad 1 \leq i, j \neq i \leq N \quad , \tag{2.10}$$

holds for sampling without replacement .

However this is only a sufficient and not a necessary condition. This condition is very strict and seldom fulfilled in practice.

Consider again the special case where the PSUs are selected with equal probabilities, i.e. simple random sampling. In Section 1.5.1 it was shown that for sampling without replacement the relation $\pi_i = nz_i$ does hold. Therefore $\hat{Y}_{pps}$ remains an unbiased estimator of the population total. By comparing (1.20) and (1.36) for sampling with and without replacement respectively, it can be seen that

$$Var_{wr}(\hat{Y}_{pps}) > Var_{wor}(\hat{Y}_{pps}) \qquad \text{since} \qquad \frac{N-n}{N-1} < 1 \quad .$$

Thus, the Kish assertion is true for single-stage simple random sampling.

2.1.3 Multistage sampling

Recall again the variances of the population total and its estimators for sampling with and without replacement as given in Sections 1.4.2 and 1.5.2 respectively:

$$Var_{wr}(\hat{Y}_{pps}) = \frac{1}{n}\sum_{i=1}^{N}\sum_{j>i}^{N} z_i z_j \left(\frac{Y_i}{z_i} - \frac{Y_j}{z_j}\right)^2 + \frac{1}{n}\sum_{i=1}^{N}\frac{Var_2(\hat{Y}_i)}{z_i} \qquad (2.11)$$

and $$\hat{Var}_{wr}(\hat{Y}_{pps}) = \frac{1}{n^2(n-1)}\sum_{i=1}^{n}\sum_{j>i}^{n}\left(\frac{\hat{Y}_i}{z_i} - \frac{\hat{Y}_j}{z_j}\right)^2 \qquad (2.12)$$

for sampling with replacement , and

$$Var_{wor}(\hat{Y}_{pps}) = \sum_{i=1}^{N}\sum_{j>i}^{N}(\pi_i\pi_j - \pi_{ij})\left(\frac{Y_i}{\pi_i} - \frac{Y_j}{\pi_j}\right)^2 + \sum_{i=1}^{N}\frac{Var_2(\hat{Y}_i)}{\pi_i} \qquad (2.13)$$

and $$\hat{Var}_{wor}(\hat{Y}_{pps}) = \sum_{i=1}^{n}\sum_{j>i}^{n}\frac{\pi_i\pi_j - \pi_{ij}}{\pi_{ij}}\left(\frac{\hat{Y}_i}{\pi_i} - \frac{\hat{Y}_j}{\pi_j}\right)^2 + \left[\frac{1}{n-1} - \frac{\pi_i\pi_j - \pi_{ij}}{\pi_{ij}}\right]\left[\frac{\hat{Var}_2(\hat{Y}_i)}{\pi_i^2} + \frac{\hat{Var}_2(\hat{Y}_j)}{\pi_j^2}\right] \qquad (2.14)$$

for sampling without replacement .

Suppose now that subsampling within the PSUs is done in the same way for sampling of PSUs with and without replacement.

If the selection probabilities, z_i, for sampling with replacement are such that they equal $1/n$ times the inclusion probabilities, π_i, for sampling without replacement , (2.11) becomes

$$Var_{wr}(\hat{Y}_{pps}) = \frac{1}{n}\sum_{i=1}^{N}\sum_{j>i}^{N}\pi_i\pi_j\left(\frac{Y_i}{\pi_i} - \frac{Y_j}{\pi_j}\right)^2 + \sum_{i=1}^{N}\frac{Var_2(\hat{Y}_i)}{\pi_i} \quad . \qquad (2.15)$$

Durbin (1953) has shown that the bias of the inappropriate variance estimator is independent of the subsampling in the PSUs and the bias remains

$$\frac{n}{n-1}\left[Var_{wr}(\hat{Y}_{pps}) - Var_{wor}(\hat{Y}_{pps})\right] \quad .$$

Note that since the second terms of (2.13) and (2.15) are identical, the difference

$$Var_{wr}(\hat{Y}_{pps}) - Var_{wor}(\hat{Y}_{pps}) \qquad (2.16)$$

is just the difference between the first-stage components of $Var_{wr}(\hat{Y}_{pps})$ and $Var_{wor}(\hat{Y}_{pps})$. Therefore the bias depends only on the first-stage variances.

It was shown in Chapter 1 that if the selection probabilities are directly proportional to the Y_i's, the first-stage variance vanishes for both sampling with replacement and sampling without replacement when $\pi_i = nz_i$. It was also shown in Chapter 1 that if a measure of size is correlated with Y_i and the selection probabilities are taken as the relative measures of size, i.e. $z_i = M_i/M_0$, the first-stage variances for both sampling with replacement and sampling without replacement , where $\pi_i = nz_i$, will be small. Thus the difference (2.16) is small and the bias will be reduced. Thus, if it is known that the first-stage component of $Var_{wr}(\hat{Y}_{pps})$ exceeds that of $Var_{wor}(\hat{Y}_{pps})$, the inappropriate variance estimator $\hat{Var}_{wr}(\hat{Y}_{pps})$ under sampling without replacement will on the average overestimate the true variance $Var_{wor}(\hat{Y}_{pps})$ to a lesser degree.
Thus, as in the case of single-stage sampling, sampling of PSUs with selection probabilities proportional to a measure of size expected to be correlated with Y, is advisable.

By comparing (2.15) for sampling with replacement and (2.13) for sampling without replacement , it can be noted that, as in the case of single-stage sampling,

$$Var_{wr}(\hat{Y}_{pps}) > Var_{wor}(\hat{Y}_{pps})$$

if the condition

$$\pi_{ij} > \pi_i \pi_j \frac{n-1}{n} \quad , \quad 1 \leq i, j \neq i \leq N \quad ,$$

holds for sampling without replacement.

Also note that if the use of $\hat{Var}_{wr}(\hat{Y}_{pps})$ under sampling without replacement leads on the average to an overestimation of $Var_{wor}(\hat{Y}_{pps})$, this overestimation decreases when multistage sampling is applied. This is due to the fact that the variance arising from the second and later stages is the same whether sampling of PSUs with or without replacement has been applied. According to Cochran (1977), Raj (1964) found in his examples that the ratio $Var_{wor}(\hat{Y}_{pps})/Var_{wr}(\hat{Y}_{pps})$ equalled about 0,79 for the first-stage variance component and 0,92 over both, the first-stage and second-and-later-stage variance components.

The Kish assertion is thus also true on the average for multistage sampling where the PSUs in the first sampling stage are selected by simple random sampling. This can also be easily verified by comparing (1.28) for sampling with replacement to (1.47) for sampling without replacement .

2.1.4 Conclusion

From the foregoing it is clear that the use of the inappropriate variance estimator $\hat{Var}_{wr}(\hat{Y}_{pps})$ will lead on the average to an overestimation of $Var_{wor}(\hat{Y}_{pps})$ if $Var_{wr}(\hat{Y}_{pps})$ exceeds $Var_{wor}(\hat{Y}_{pps})$. This is true for single-stage as well as multistage sampling. However, the relation $Var_{wr}(\hat{Y}_{pps}) > Var_{wor}(\hat{Y}_{pps})$ cannot be proved in the general case since it depends on the particular sampling scheme applied. Various sampling schemes without replacement for which $\pi_i = nz_i$ have been proposed by various authors. Some of these sampling schemes will be considered in Section 2.2. For all the different sampling procedures considered, it was shown that $Var_{wr}(\hat{Y}_{pps}) > Var_{wor}(\hat{Y}_{pps})$ and the Kish assertion is thus true on the average for these procedures.

2.2 SAMPLING PROCEDURES FOR WHICH $\pi_i = nz_i$

2.2.1 Introduction

According to Sampford(1967), a sample of n different PSUs, where PSU i has probability $\pi_i = nz_i$ of being included in the sample, can be selected in practice in three ways:

(i) For each possible sample of size n, the required probabilities $\pi_i = nz_i$ are calculated and one sample is selected according to these probabilities.

(ii) The PSUs are selected strictly without replacement where the drawing probabilities, p_{i_m}, for each draw are chosen in such way that equation (1.9) equals nz_i. Note that in this case the drawing probabilities are based on the initial selection probabilities.

(iii) The PSUs are selected using a rejective procedure where the drawing probabilities are altered after the first draw in order to ensure that $\pi_i = nz_i$.

Method (i) is impracticable, especially if a stratum contains many PSUs. Various sampling schemes that make use of methods (ii) and (iii) will now be discussed. The first two of these procedures are suitable for samples of size two and will be considered in Section 2.2.2. The latter two procedures are also suitable for sample sizes greater than two and will be considered in Section 2.2.3. For each procedure an example will be given and it can be noted that for each sample size the various procedures lead to the same sets of inclusion and joint inclusion probabilities.

In Section 2.2.4 artificial populations are used to investigate the degree of overestimation of $Var_{wr}(\hat{Y}_{pps})$ and the resulting bias. The individual values of $\hat{Var}_{wr}(\hat{Y}_{pps})$ will also be compared to those of $\hat{Var}_{wor}(\hat{Y}_{pps})$.

2.2.2 Procedures suitable for $n = 2$

The two sampling procedures considered here are due to Brewer (1963) and Durbin (1967) respectively. These procedures make use of method (ii) of Section 2.2.1. The example of Section 1.3 is used to illustrate the two sampling schemes.

(a) Brewer's Procedure

For a sample of size 2, the first PSU is selected with drawing probability proportional to $z_i(1-z_i)/(1-2z_i)$ and the second with drawing probability $z_j/(1-z_i)$, where PSU i is selected first and j second.

The drawing probabilities for the first and second draw are thus given by

$$p_{i_1} = \frac{z_i(1-z_i)}{D(1-2z_i)} \tag{2.17}$$

$$\text{and} \qquad p_{j_2} = \frac{z_j}{1-z_i} \quad , \tag{2.18}$$

$$\text{where} \qquad D = \frac{1}{2}\left(1 + \sum_{i=1}^{N} \frac{z_i}{1-2z_i}\right) \quad .$$

By using (1.9) it is clear that the inclusion probabilities are given by

$$\begin{aligned}
\pi_i &= \delta_i(1) + \delta_i(2) \\
&= \frac{z_i(1-z_i)}{D(1-2z_i)} + \sum_{j \neq i}^{N} \frac{z_j(1-z_j)}{D(1-2z_j)} \frac{z_i}{(1-z_j)} \\
&= \frac{z_i}{D}\left[\frac{(1-z_i)}{(1-2z_i)} + \sum_{j \neq i}^{N} \frac{z_j}{(1-2z_j)}\right] \\
&= \frac{z_i}{D}\left[1 + \sum_{j=1}^{N} \frac{z_j}{(1-2z_j)}\right] \\
&= 2z_i \quad .
\end{aligned} \tag{2.19}$$

Using (1.11), the joint inclusion probabilities are given by

$$\begin{aligned}
\pi_{ij} &= \delta_{ij}(1,2) + \delta_{ij}(2,1) \\
&= \frac{z_i(1-z_i)}{D(1-2z_i)} \frac{z_j}{(1-z_i)} + \frac{z_j(1-z_j)}{D(1-2z_j)} \frac{z_i}{(1-z_j)} \\
&= \frac{z_i z_j}{D}\left(\frac{1}{1-2z_i} + \frac{1}{1-2z_j}\right) \quad .
\end{aligned} \tag{2.20}$$

According to Cochran (1977), Rao (1965) has shown that for Brewer's procedure, $(\pi_i\pi_j - \pi_{ij}) > 0$ for all $i \neq j$, so that the variance estimate obtained is always positive. Brewer (1963) has shown that the variance of $\hat{Y}_{pps}$ obtained by his method is always less than that obtained by sampling with replacement .

Sampford (1967) has extended Brewer's method for the case where $n > 2$. This procedure will be discussed in Section 2.2.3.

Example

One of the following 12 possible samples can be selected from the universe given on page 7, Section 1.3:

Sample no.	*Sample*	*Drawing Probabilities*		$P(s_n)$
	$s_n = (i,j)$	p_{i_1}	p_{j_2}	$p_{i_1} \cdot p_{j_2}$
1	1, 2	0, 1267	0, 5000	0, 0634
2	1, 3	0, 1267	0, 3750	0, 0475
3	1, 4	0, 1267	0, 1250	0, 0158
4	2, 1	0, 5703	0, 3333	0, 1901
5	2, 3	0, 5703	0, 5000	0, 2851
6	2, 4	0, 5703	0, 1667	0, 0951
7	3, 1	0, 2495	0, 2857	0, 0713
8	3, 2	0, 2495	0, 5714	0, 1426
9	3, 4	0, 2495	0, 1429	0, 0356
10	4, 1	0, 0535	0, 2222	0, 0119
11	4, 2	0, 0535	0, 4444	0, 0238
12	4, 3	0, 0535	0, 3333	0, 0178
			Total	1, 0000

where $D = 2, 1042$.

From the table it can be seen that the inclusion probabilities comply with the requirement $\pi_i = nz_i$. For instance,

$$\begin{aligned}\pi_1 &= \sum_{s_n \ni 1} P(s_n) \\ &= P(1,2) + P(1,3) + P(1,4) + P(2,1) + P(3,1) + P(4,1) \\ &= 0, 4 = 2z_1 \quad .\end{aligned}$$

It can also be seen that the joint inclusion probabilities obtained by the example comply with (2.20). For instance,

$$\pi_{12} = \sum_{s_n \ni 1,2} P(s_n) = P(1,2) + P(2,1) = 0, 2535$$

and by using (2.20),

$$\pi_{12} = \frac{z_1 z_2}{D}\left(\frac{1}{1-2z_1} + \frac{1}{1-2z_2}\right) = 0,2535 \quad .$$

The following joint inclusion probabilities are obtained:

$$\pi_{12} = 0,2535 \quad , \quad \pi_{13} = 0,1188 \quad , \quad \pi_{14} = 0,0277 \quad ,$$
$$\pi_{23} = 0,4277 \quad , \quad \pi_{24} = 0,1189 \quad , \quad \pi_{34} = 0,0534 \quad .$$

(b) Durbin's Procedure

Durbin selects the first PSU with drawing probability z_i and the second with drawing probability proportional to $z_i\left[(1-2z_i)^{-1} + (1-2z_j)^{-1}\right]$, where PSU i is selected first and j second.

The drawing probabilities for the first and second draw are thus given by

$$p_{i_1} = z_i \tag{2.21}$$

$$\text{and} \quad p_{j_2} = \frac{z_j}{D}\left(\frac{1}{1-2z_i} + \frac{1}{1-2z_j}\right) \quad , \tag{2.22}$$

$$\text{where} \quad D = 1 + \sum_{j=1}^{N} \frac{z_j}{1-2z_j} \quad .$$

By using (1.9) it is clear that the inclusion probabilities are given by

$$\begin{aligned}
\pi_i &= \delta_i(1) + \delta_i(2) \\
&= z_i + \sum_{j \neq i}^{N} \frac{z_j}{D}\left(\frac{1}{1-2z_i} + \frac{1}{1-2z_j}\right) z_i \\
&= z_i + \frac{z_i}{D}\left[\frac{1}{(1-2z_i)}\left(\sum_{i=1}^{N} z_j - z_i\right) + \sum_{j=1}^{N} \frac{z_j}{(1-2z_j)} - \frac{z_i}{(1-2z_i)}\right] \\
&= z_i + \frac{z_i}{D}\left[1 + \sum_{j=1}^{N} \frac{z_j}{(1-2z_j)}\right] \\
&= 2z_i \quad .
\end{aligned} \tag{2.23}$$

Using (1.11), the joint inclusion probabilities are given by

$$\begin{aligned}
\pi_{ij} &= \delta_{ij}(1,2) + \delta_{ij}(2,1) \\
&= z_i \frac{z_j}{D}\left(\frac{1}{1-2z_i} + \frac{1}{1-2z_j}\right) + z_j \frac{z_i}{D}\left(\frac{1}{1-2z_j} + \frac{1}{1-2z_i}\right) \\
&= \frac{2z_i z_j}{D}\left(\frac{1}{1-2z_i} + \frac{1}{1-2z_j}\right) \quad .
\end{aligned} \tag{2.24}$$

It can be seen that π_{ij} obtained by (2.24) is identical to (2.20) obtained by Brewer's method. Brewer and Hanif (1983) state that Durbin's procedure is not generally applicable for $n > 2$. Durbin (1967) has shown that the factor $(\pi_i\pi_j - \pi_{ij})/\pi_{ij} > 0$, so that the variance estimate obtained by Durbin's method is always positive. Sutradhar (1984) gives a simple proof that the variance obtained by Durbin's procedure is always less than that obtained by sampling with replacement .

Example

One of the following 12 possible samples can be selected from the universe given on page 7, Section 1.3:

Sample no.	*Sample*	*Drawing Probabilities*		$P(s_n)$
	$s_n = (i, j)$	p_{i_1}	p_{j_2}	$p_{i_1} \cdot p_{j_2}$
1	1, 2	0, 2	0, 6337	0, 1267
2	1, 3	0, 2	0, 2970	0, 0594
3	1, 4	0, 2	0, 0693	0, 0139
4	2, 1	0, 4	0, 3168	0, 1267
5	2, 3	0, 4	0, 5347	0, 2139
6	2, 4	0, 4	0, 1485	0, 0594
7	3, 1	0, 3	0, 1980	0, 0594
8	3, 2	0, 3	0, 7129	0, 2139
9	3, 4	0, 3	0, 0891	0, 0267
10	4, 1	0, 1	0, 1386	0, 0139
11	4, 2	0, 1	0, 5941	0, 0594
12	4, 3	0, 1	0, 2673	0, 0267
			Total	1, 0000

where $D = 4, 2083$.

From the table it can be seen that the inclusion probabilities comply with the requirement $\pi_i = nz_i$. For instance,

$$\begin{aligned}\pi_1 &= \sum_{s_n \ni 1} P(s_n) \\ &= P(1,2) + P(1,3) + P(1,4) + P(2,1) + P(3,1) + P(4,1) \\ &= 0, 4 = 2z_1 \quad .\end{aligned}$$

It can also be seen that the joint inclusion probabilities obtained by the example comply with (2.24). For instance,

$$\pi_{12} = \sum_{s_n \ni 1,2} P(s_n) = P(1,2) + P(2,1) = 0, 2534$$

and by using (2.24),

$$\pi_{12} = \frac{2z_1z_2}{D}\left(\frac{1}{1-2z_1}+\frac{1}{1-2z_2}\right) = 0,2535 \quad .$$

The joint inclusion probabilities obtained are the same as those obtained by Brewer's procedure.

Although Durbin's procedure leads to the same sets of inclusion and joint inclusion probabilities as Brewer's procedure, it can be seen that Durbin's procedure is symmetrical in the sense that $P(i,j) = P(j,i)$, i.e. the probability of selecting a sample which contains PSUs i and j does not depend on the order of draw of PSUs i and j. For Brewer's procedure, if $z_i > z_j$ then $P(i,j) > P(j,i)$, i.e. if the order of draw is taken into account, the probability of selecting a sample which contains both PSUs i and j does depend on which PSU was selected first. However the order of draw is generally not of importance. We are only concerned about the probability that a sample selected contains both PSUs i and j, i.e π_{ij}.

Now consider condition (2.10):

(i,j)	π_{ij}	$\frac{1}{2}\pi_i\pi_j$
1,2	0,2535	0,16
1,3	0,1188	0,12
1,4	0,0277	0,04
2,3	0,4277	0,24
2,4	0,1189	0,08
3,4	0,0534	0,06

Although $Var_{wr}(\hat{Y}_{pps})$ exceeds $Var_{wor}(\hat{Y}_{pps})$ it can be seen that for 3 of the 6 possible cases relation (2.10) does not hold. It can be noted that this happens whenever a sample of two PSUs having low selection probabilities is selected. The total probability that a sample is selected for which relation (2.10) does not hold is in this case 0,1999.

In practice, Durbin's procedure is probably preferred to Brewer's procedure since the drawing probabilities for the first are easier to calculate than those for the latter.

2.2.3 Procedures suitable for $n > 2$

The first procedure considered here is a generalization of the one due to Brewer (1963) discussed in Section 2.2.2. It is due to Sampford (1967) and is called the Durbin-Sampford procedure. This procedure makes use of method (ii) of Section 2.2.1. The second procedure considered here is a rejective procedure, i.e. it makes use of method (iii) of Section 2.2.1. It is also due to Sampford (1967).

Note that most of the descriptions used in this paragraph are reproduced from Sampford (1967) with appropriate changes in notation.

For both procedures, define for each unit

$$\lambda_i = \frac{z_i}{1 - nz_i} \quad . \tag{2.25}$$

Also, denote a sample of $m \leq N$ different PSUs $i_1, i_2, \ldots, i_m$ by s_m and define L_m by

$$L_0 = 1 \quad , \quad L_m = \sum_{s_m} \lambda_{i_1} \lambda_{i_2} \ldots \lambda_{i_m} \qquad (1 \leq m \leq N) \quad , \tag{2.26}$$

where summation $\sum_{s_m}$ extends over all possible samples that contain m different PSUs.

The probability that a particular sample s_n, containing n different PSUs, is selected is now given by

$$P(s_n) = K_n \sum_{u=1}^{n} z_{i_u} \prod_{v \neq u}^{n} \lambda_{i_v}$$

$$\text{or by} \quad P(s_n) = nK_n \lambda_{i_1} \lambda_{i_2} \ldots \lambda_{i_n} \left(1 - \sum_{u=1}^{n} z_{i_u}\right) \quad , \tag{2.27}$$

where K_n is a constant multiplier and is given by

$$K_n = \left(\sum_{t=1}^{n} \frac{t L_{n-t}}{n^t}\right)^{-1} \quad . \tag{2.28}$$

For the example used to illustrate the two procedures, a sample of size 3 is selected from the following universe:

PSU, i	1	2	3	4	5
MOS, M_i	15	20	25	30	10
$z_i = M_i/M_0$	0,15	0,2	0,25	0,3	0,1
$\lambda_i = z_i/(1-3z_i)$	0,2727	0,5	1	3	0,1429

(a) Durbin-Sampford's Procedure

Since it is rather difficult to explain this procedure for the general n it will be only discussed for the case $n = 3$.

The first PSU, say the i-th, is selected with drawing probability

$$p_{i_1} = c_1 \lambda_i \sum_{j \neq i}^{N} \lambda_j (1 - z_i - z_j) \tag{2.29a}$$

$$= c_1 \lambda_i c_2^{-1} \quad , \tag{2.29b}$$

the second, say the j-th, with drawing probability

$$p_{j_2} = c_2 \lambda_j (1 - z_i - z_j) \tag{2.30a}$$

$$= c_2 \lambda_j c_3^{-1} \tag{2.30b}$$

and the third, say the k-th, with drawing probability

$$p_{k_3} = c_3 z_k \quad . \tag{2.31}$$

By using (1.5), the probability of selecting a sample s_3, containing PSUs i, j and k in this order, is given by

$$P(s_3) = p_{i_1} p_{j_2} p_{k_3} = c_1 \lambda_i \lambda_j z_k \quad .$$

Thus, if the order of draw is ignored, the probability of selecting a sample containing PSUs i, j and k is given by

$$\begin{aligned} P(s_3) =& 2c_1\lambda_i\lambda_j z_k + 2c_1\lambda_i\lambda_k z_j + 2c_1\lambda_j\lambda_k z_i \\ =& 2c_1\lambda_i\lambda_j\lambda_k \left(\frac{z_k}{\lambda_k} + \frac{z_j}{\lambda_j} + \frac{z_i}{\lambda_i}\right) \\ =& 2c_1\lambda_i\lambda_j\lambda_k \left|(1-3z_k) + (1-3z_j) + (1-3z_i)\right| \\ =& 3.2c_1\lambda_i\lambda_j\lambda_k \left(1 - \sum_{u=1}^{n} z_{i_u}\right) \end{aligned} \tag{2.32}$$

and it can be seen that this probability complies with (2.27) if $c_1 = \frac{1}{2}K_3$.

Sampford (1967) has shown that the inclusion probabilities are given by $\pi_i = nz_i$ for any n. Since this proof is rather complicated it will not be given here.

The formula for the calculation of the joint inclusion probabilities as given by Sampford (1967) is rather complicated for the general n. Therefore it will now be derived for the case where $n = 3$. Note that for $n = 3$, the term $\delta_{ij}(1,2)$, for example, used in the derivation denotes the probability that PSU i is drawn first, PSU j second and any other PSU, except i and j, third.

Using (1.11), the joint inclusion probabilities are given by

$$\begin{aligned} \pi_{ij} =& \delta_{ij}(1,2) + \delta_{ij}(1,3) + \delta_{ij}(2,1) + \delta_{ij}(2,3) + \delta_{ij}(3,1) + \delta_{ij}(3,2) \\ =& c_1\lambda_i c_2^{-1} c_2 \lambda_j c_3^{-1} c_3 \sum_{k=1(\neq(i\neq j))}^{N} z_k + c_1\lambda_i c_2^{-1} c_2 \sum_{k=1(\neq(i\neq j))}^{N} \lambda_k c_3^{-1} c_3 z_j \\ &+ c_1\lambda_j c_2^{-1} c_2 \lambda_i c_3^{-1} c_3 \sum_{k=1(\neq(i\neq j))}^{N} z_k + c_1 \sum_{k=1(\neq(i\neq j))}^{N} \lambda_k c_2^{-1} c_2 \lambda_i c_3^{-1} c_3 z_j \\ &+ c_1\lambda_j c_2^{-1} c_2 \sum_{k=1(\neq(i\neq j))}^{N} \lambda_k c_3^{-1} c_3 z_i + c_1 \sum_{k=1(\neq(i\neq j))}^{N} \lambda_k c_2^{-1} c_2 \lambda_j c_3^{-1} c_3 z_i \\ =& 2c_1 \left(\lambda_i\lambda_j \sum_{k=1(\neq(i\neq j))}^{N} z_k + \lambda_i z_j \sum_{k=1(\neq(i\neq j))}^{N} \lambda_k + \lambda_j z_i \sum_{k=1(\neq(i\neq j))}^{N} \lambda_k \right) \\ =& 2c_1 \left(\lambda_i\lambda_j(1 - z_i - z_j) + (\lambda_i z_j + \lambda_j z_i) \sum_{k=1(\neq(i\neq j))}^{N} \lambda_k \right) \quad . \end{aligned} \tag{2.33}$$

Cochran (1977) states that the variance estimate is always positive for this method.

Example

For the universe of the example given earlier in this section on page 38

$$L_0 = 1 \quad , \quad L_1 = \sum_{i_1} \lambda_{i_1} = 4,9156 \quad \text{and} \quad L_2 = \sum_{i_1 i_2} \lambda_{i_1} \lambda_{i_2} = 6,90917 \quad ,$$

such that

$$K_3 = \left(\frac{L_2}{3} + \frac{2L_1}{9} + \frac{3L_0}{27} \right)^{-1} = 0,2851827$$

and $$c_1 = \frac{K_3}{2} = 0,1425913 \quad .$$

One of the following 60 samples can be selected:

Sample no.	*Sample*	*Drawing*		*Probabilities*	$P(s_n)$
	$s_n = (i,j,k)$	p_{i_1}	p_{j_2}	p_{k_3}	$p_{i_1} \cdot p_{j_2} \cdot p_{k_3}$
1	1, 2, 3	0, 1042954	0, 1211703	0, 3846153	0, 00486057
2	1, 3, 2	0, 1042954	0, 2236990	0, 3333333	0, 00777692
3	2, 1, 3	0, 1873139	0, 0674669	0, 3846153	0, 00486057
4	2, 3, 1	0, 1873139	0, 2093415	0, 2727272	0, 01069430
5	3, 1, 2	0, 2643977	0, 0882412	0, 3333333	0, 00777692
6	3, 2, 1	0, 2643977	0, 1483091	0, 2727272	0, 01069430
				$\pi_{123} =$	0, 0466634
7	1, 2, 4	0, 1042954	0, 1211703	0, 4615384	0, 00583269
8	1, 4, 2	0, 1042954	0, 6151723	0, 3636363	0, 02333070
9	2, 1, 4	0, 1873139	0, 0674669	0, 4615384	0, 00583269
10	2, 4, 1	0, 1873139	0, 5709315	0, 3000000	0, 03208300
11	4, 1, 2	0, 4002787	0, 1602874	0, 3636363	0, 02333070
12	4, 2, 1	0, 4002787	0, 2671725	0, 3000000	0, 03208300
				$\pi_{124} =$	0, 1224926
13	1, 2, 5	0, 1042954	0, 1211703	0, 1538461	0, 00194423
14	1, 5, 2	0, 1042954	0, 0399582	0, 2666666	0, 00111132
15	2, 1, 5	0, 1873139	0, 0674669	0, 1538461	0, 00194422
16	2, 5, 1	0, 1873139	0, 0380735	0, 2142857	0, 00152822
17	5, 1, 2	0, 0612210	0, 0680723	0, 2666666	0, 00111132
18	5, 2, 1	0, 0612210	0, 1164909	0, 2142857	0, 00152821
				$\pi_{125} =$	0, 00916742

Sample no.	*Sample*	*Drawing*		*Probabilities*	$P(s_n)$
	$s_n = (i, j, k)$	p_{i_1}	p_{j_2}	p_{k_3}	$p_{i_1} \cdot p_{j_2} \cdot p_{k_3}$
19	1, 3, 4	0, 1042954	0, 2236990	0, 4999999	0, 01166530
20	1, 4, 3	0, 1042954	0, 6151723	0, 4545454	0, 02916340
21	3, 1, 4	0, 2643977	0, 0882412	0, 4999999	0, 01166530
22	3, 4, 1	0, 2643977	0, 7280629	0, 3333333	0, 06416600
23	4, 1, 3	0, 4002787	0, 1602874	0, 4545454	0, 02916340
24	4, 3, 1	0, 4002787	0, 4809105	0, 3333333	0, 06416600
				$\pi_{134} =$	0, 2099894
25	1, 3, 5	0, 1042954	0, 2236990	0, 1666666	0, 00388846
26	1, 5, 3	0, 1042954	0, 0399582	0, 3333333	0, 00138915
27	3, 1, 5	0, 2643977	0, 0882412	0, 1666666	0, 00388845
28	3, 5, 1	0, 2643977	0, 0500934	0, 2307692	0, 00305644
29	5, 1, 3	0, 0612210	0, 0680723	0, 3333333	0, 00138915
30	5, 3, 1	0, 0612210	0, 2163403	0, 2307692	0, 00305643
				$\pi_{135} =$	0, 0166678
31	1, 4, 5	0, 1042954	0, 6151723	0, 1818181	0, 01166530
32	1, 5, 4	0, 1042954	0, 0399582	0, 3999999	0, 00166698
33	4, 1, 5	0, 4002787	0, 1602874	0, 1818181	0, 01166530
34	4, 5, 1	0, 4002787	0, 0916294	0, 2499999	0, 00916932
35	5, 1, 4	0, 0612210	0, 0680723	0, 3999999	0, 00166698
36	5, 4, 1	0, 0612210	0, 5990963	0, 2499999	0, 00916931
				$\pi_{145} =$	0, 0450030
37	2, 3, 4	0, 1873139	0, 2093415	0, 5454545	0, 02138860
38	2, 4, 3	0, 1873139	0, 5709315	0, 5000000	0, 05347170
39	3, 2, 4	0, 2643977	0, 1483091	0, 5454545	0, 02138860
40	3, 4, 2	0, 2643977	0, 7280629	0, 4444444	0, 08555470
41	4, 2, 3	0, 4002787	0, 2671725	0, 5000000	0, 05347170
42	4, 3, 2	0, 4002787	0, 4809105	0, 4444444	0, 08555470
				$\pi_{234} =$	0, 3208300
43	2, 3, 5	0, 1873139	0, 2093415	0, 1818181	0, 00712955
44	2, 5, 3	0, 1873139	0, 0380735	0, 3571428	0, 00254703
45	3, 2, 5	0, 2643977	0, 1483091	0, 1818181	0, 00712956
46	3, 5, 2	0, 2643977	0, 0500934	0, 3076923	0, 00407525
47	5, 2, 3	0, 0612210	0, 1164909	0, 3571428	0, 00254703
48	5, 3, 2	0, 0612210	0, 2163403	0, 3076923	0, 00407525
				$\pi_{235} =$	0, 0275034

Sample no.	*Sample*	*Drawing*		*Probabilities*	$P(s_n)$
	$s_n = (i,j,k)$	p_{i_1}	p_{j_2}	p_{k_3}	$p_{i_1} \cdot p_{j_2} \cdot p_{k_3}$
49	2, 4, 5	0, 1873139	0, 5709315	0, 2000000	0, 02138860
50	2, 5, 4	0, 1873139	0, 0380735	0, 4285714	0, 00305644
51	4, 2, 5	0, 4002787	0, 2671725	0, 2000000	0, 02138860
52	4, 5, 2	0, 4002787	0, 0916294	0, 3333333	0, 01222570
53	5, 2, 4	0, 0612210	0, 1164909	0, 4285714	0, 00305643
54	5, 4, 2	0, 0612210	0, 5990963	0, 3333333	0, 01222570
				$\pi_{245} =$	0, 0733414
55	3, 4, 5	0, 2643977	0, 7280629	0, 2222222	0, 04277730
56	3, 5, 4	0, 2643977	0, 0500934	0, 4615384	0, 00611288
57	4, 3, 5	0, 4002787	0, 4809105	0, 2222222	0, 04277730
58	4, 5, 3	0, 4002787	0, 0916294	0, 4166666	0, 01528220
59	5, 3, 4	0, 0612210	0, 2163403	0, 4615384	0, 00611287
60	5, 4, 3	0, 0612210	0, 5990963	0, 4166666	0, 01528210
				$\pi_{345} =$	0, 1283445
				Total	1, 0000029

From the table it can also be seen that the inclusion probabilities comply with the requirement $\pi_i = nz_i$. For instance,

$$\begin{aligned}\pi_1 &= \sum_{s_n \ni 1} P(s_n) \\ &= \pi_{123} + \pi_{124} + \pi_{125} + \pi_{134} + \pi_{135} + \pi_{145} \\ &= 0,45 = 3z_1 \quad .\end{aligned}$$

It can also be seen that the joint inclusion probabilities obtained by the example comply with (2.33). For instance,

$$\pi_{12} = \sum_{s_n \ni 1,2} P(s_n) \quad = \pi_{123} + \pi_{124} + \pi_{125} \quad = 0, 1783234$$

and by using (2.33),

$$\pi_{12} = 2c_1 \left(\lambda_1 \lambda_2 (1 - z_1 - z_2) + (\lambda_1 z_2 + \lambda_2 z_1) \sum_{k=3}^{5} \lambda_k \right) = 0, 1783243 \quad .$$

The following joint inclusion probabilities are obtained:

$\pi_{12} = 0, 1783234$, $\pi_{13} = 0, 2733206$, $\pi_{14} = 0, 377485$,

$\pi_{15} = 0, 0708382$, $\pi_{23} = 0, 3949968$, $\pi_{24} = 0, 516664$,

$\pi_{25} = 0, 1100122$, $\pi_{34} = 0, 6591639$, $\pi_{35} = 0, 1725157$, $\pi_{45} = 0, 2466889$.

(b) Sampford's Rejective Procedure

For a sample of n PSUs, the first PSU is selected with selection probability z_i and the PSU drawn is replaced. At each subsequent draw a PSU is selected with replacement with selection probability

$$\lambda_i \Big/ \left(\sum_{i=1}^{N} \lambda_i\right) \quad .$$

For this method,

$$P(s_n) = (n-1)!^{-1} K_n \left(\sum_{i=1}^{N} \lambda_i\right)^{n-1} \times \text{coefficient of } t_{i_1} \ldots t_{i_n} \text{ in } W(t) \quad ,$$

$$\text{where} \quad W(t) = \left(\sum_{i=1}^{N} z_i t_i\right) \left(\sum_{i=1}^{N} \lambda_i t_i \Big/ \sum_{i=1}^{N} \lambda_i\right)^{n-1} \tag{2.34}$$

is the probability generating function for sampling with replacement if the PSUs are drawn as mentioned above. By this scheme the probability of obtaining a sample without repetitions is proportional to probability (2.27). Probability (2.27) can be achieved if a rejective scheme is applied. This means, the sample of n PSUs is selected with replacement where the first drawing is made with probability z_i and the other $(n-1)$ drawings with probabilities $\lambda_i/(\sum \lambda_i)$. The sample is accepted if it contains n different PSUs, otherwise it is rejected and a new sample is selected. This process is continued until an acceptable sample is obtained. Note that in practice a sample can be rejected as soon as the first duplicate appears in the sample.

The probability of obtaining an acceptable sample at any trial is given by

$$P = (n-1)! \Big/ \left[K_n \left(\sum_{i=1}^{N} \lambda_i\right)^{n-1}\right] \quad . \tag{2.35}$$

Thus, the expected number of samples that have to be drawn in order to obtain an acceptable sample is given by the reciprocal of (2.35).

Sampford has shown that the inclusion probabilities comply with the requirement $\pi_i = nz_i$. Since $\pi_i = nz_i$,

$$\sum_{i=1}^{N} \pi_i = n$$

and the first relation given in (1.13) is thus satisfied.

The formula for the calculation of the joint inclusion probabilities is rather complex for the general n. Therefore it will be derived here only for $n = 3$. By using (1.11) π_{ij} is given by

$$\begin{aligned}
\pi_{ij} =& \delta_{ij}(1,2) + \delta_{ij}(1,3) + \delta_{ij}(2,1) + \delta_{ij}(2,3) + \delta_{ij}(3,1) + \delta_{ij}(3,2) \\
=& \frac{1}{P}\left[2z_i.(\lambda_j) \Big/ \left(\sum_{i=1}^{N}\lambda_i\right) . \left(\sum_{k=1(\neq(i\neq j))}^{N}\lambda_k\right) \Big/ \left(\sum_{i=1}^{N}\lambda_i\right)\right. \\
&+ \quad 2z_j.(\lambda_i) \Big/ \left(\sum_{i=1}^{N}\lambda_i\right) . \left(\sum_{k=1(\neq(i\neq j))}^{N}\lambda_k\right) \Big/ \left(\sum_{i=1}^{N}\lambda_i\right) \\
&\left. + \quad 2 \sum_{k=1(\neq(i\neq j))}^{N} z_k.(\lambda_i) \Big/ \left(\sum_{i=1}^{N}\lambda_i\right) . (\lambda_j) \Big/ \left(\sum_{i=1}^{N}\lambda_i\right)\right] \\
=& \frac{2}{P}\left(\sum_{i=1}^{N}\lambda_i\right)^{-2} \left(\lambda_i\lambda_j(1 - z_i - z_j) + (\lambda_i z_j + \lambda_j z_i) \sum_{k=1(\neq(i\neq j))}^{N} \lambda_k \right) \qquad . \quad (2.36)
\end{aligned}$$

It can be seen that π_{ij} obtained by (2.36) is identical to (2.33) obtained by the Durbin-Sampford method, since

$$\frac{1}{P}\left(\sum_{i=1}^{N}\lambda_i\right)^{-2} = \frac{1}{2}K_3 = c_1 \qquad .$$

Therefore the latter two relations given in (1.13) also hold for Sampford's rejective procedure.

According to Cochran (1977), the variance estimate obtained by this rejective method is always positive and Gabler (1981) has shown that the variance is always less than that obtained by sampling with replacement . Since the Durbin-Sampford procedure produces the same sets of inclusion and joint inclusion probabilities as Sampford's rejective procedure, this result also holds for the Durbin-Sampford procedure.

Example

For the universe of the example given on page 38 earlier in this section, the value of K_3 is the same as that for the Durbin-Sampford procedure. Thus, the probability of obtaining an acceptable sample at any trial is given by $P = 0,2902376$. One of the following 60 samples can be selected:

Sample no.	*Sample*	*Drawing*		*Probabilities*	$P(s_n)$
	$s_n = (i, j, k)$	p_{i_1}	p_{j_2}	p_{k_3}	$p_{i_1} \cdot p_{j_2} \cdot p_{k_3} \cdot P$
1	1, 2, 3	0, 15	0, 1017169	0, 2034339	0, 01069430
2	1, 3, 2	0, 15	0, 2034339	0, 1017169	0, 01069430
3	2, 1, 3	0, 20	0, 0554764	0, 2034339	0, 00777692
4	2, 3, 1	0, 20	0, 2034339	0, 0554764	0, 00777692
5	3, 1, 2	0, 25	0, 0554764	0, 1017169	0, 00486057
6	3, 2, 1	0, 25	0, 1017169	0, 0554764	0, 00486057
				$\pi_{123} =$	0, 0466634
7	1, 2, 4	0, 15	0, 1017169	0, 6103018	0, 03208300
8	1, 4, 2	0, 15	0, 6103018	0, 1017169	0, 03208300
9	2, 1, 4	0, 20	0, 0554764	0, 6103018	0, 02333070
10	2, 4, 1	0, 20	0, 6103018	0, 0554764	0, 02333070
11	4, 1, 2	0, 30	0, 0554764	0, 1017169	0, 00583268
12	4, 2, 1	0, 30	0, 1017169	0, 0554764	0, 00583268
				$\pi_{124} =$	0, 1224926
13	1, 2, 5	0, 15	0, 1017169	0, 0290707	0, 00152822
14	1, 5, 2	0, 15	0, 0290707	0, 1017169	0, 00152822
15	2, 1, 5	0, 20	0, 0554764	0, 0290707	0, 00111132
16	2, 5, 1	0, 20	0, 0290707	0, 0554764	0, 00111132
17	5, 1, 2	0, 10	0, 0554764	0, 1017169	0, 00194423
18	5, 2, 1	0, 10	0, 1017169	0, 0554764	0, 00194423
				$\pi_{125} =$	0, 0091674
19	1, 3, 4	0, 15	0, 2034339	0, 6103018	0, 06416600
20	1, 4, 3	0, 15	0, 6103018	0, 2034339	0, 06416600
21	3, 1, 4	0, 25	0, 0554764	0, 6103018	0, 02916340
22	3, 4, 1	0, 25	0, 6103018	0, 0554764	0, 02916340
23	4, 1, 3	0, 30	0, 0554764	0, 2034339	0, 01166530
24	4, 3, 1	0, 30	0, 2034339	0, 0554764	0, 01166530
				$\pi_{134} =$	0, 2099894
25	1, 3, 5	0, 15	0, 2034339	0, 0290707	0, 00305644
26	1, 5, 3	0, 15	0, 0290707	0, 2034339	0, 00305644
27	3, 1, 5	0, 25	0, 0554764	0, 0290707	0, 00138915
28	3, 5, 1	0, 25	0, 0290707	0, 0554764	0, 00138915
29	5, 1, 3	0, 10	0, 0554764	0, 2034339	0, 00388846
30	5, 3, 1	0, 10	0, 2034339	0, 0554764	0, 00388846
				$\pi_{135} =$	0, 0166678

Sample no.	Sample	Drawing		Probabilities	$P(s_n)$
	$s_n = (i, j, k)$	p_{i_1}	p_{j_2}	p_{k_3}	$p_{i_1} \cdot p_{j_2} \cdot p_{k_3} \cdot P$
31	1, 4, 5	0, 15	0, 6103018	0, 0290707	0, 00916933
32	1, 5, 4	0, 15	0, 0290707	0, 6103018	0, 00916933
33	4, 1, 5	0, 30	0, 0554764	0, 0290707	0, 00166698
34	4, 5, 1	0, 30	0, 0290707	0, 0554764	0, 00166698
35	5, 1, 4	0, 10	0, 0554764	0, 6103018	0, 01166530
36	5, 4, 1	0, 10	0, 6103018	0, 0554764	0, 01166530
				$\pi_{145} =$	0, 0450030
37	2, 3, 4	0, 20	0, 2034339	0, 6103018	0, 08555470
38	2, 4, 3	0, 20	0, 6103018	0, 2034339	0, 08555470
39	3, 2, 4	0, 25	0, 1017169	0, 6103018	0, 05347170
40	3, 4, 2	0, 25	0, 6103018	0, 1017169	0, 05347170
41	4, 2, 3	0, 30	0, 1017169	0, 2034339	0, 02138860
42	4, 3, 2	0, 30	0, 2034339	0, 1017169	0, 02138860
				$\pi_{234} =$	0, 3208300
43	2, 3, 5	0, 20	0, 2034339	0, 0290707	0, 00407525
44	2, 5, 3	0, 20	0, 0290707	0, 2034339	0, 00407525
45	3, 2, 5	0, 25	0, 1017169	0, 0290707	0, 00254703
46	3, 5, 2	0, 25	0, 0290707	0, 1017169	0, 00254703
47	5, 2, 3	0, 10	0, 1017169	0, 2034339	0, 00712955
48	5, 3, 2	0, 10	0, 2034339	0, 1017169	0, 00712955
				$\pi_{235} =$	0, 0275034
49	2, 4, 5	0, 20	0, 6103018	0, 0290707	0, 01222570
50	2, 5, 4	0, 20	0, 0290707	0, 6103018	0, 01222570
51	4, 2, 5	0, 30	0, 1017169	0, 0290707	0, 00305644
52	4, 5, 2	0, 30	0, 0290707	0, 1017169	0, 00305644
53	5, 2, 4	0, 10	0, 1017169	0, 6103018	0, 02138860
54	5, 4, 2	0, 10	0, 6103018	0, 1017169	0, 02138860
				$\pi_{245} =$	0, 0733414
55	3, 4, 5	0, 25	0, 6103018	0, 0290707	0, 01528220
56	3, 5, 4	0, 25	0, 0290707	0, 6103018	0, 01528220
57	4, 3, 5	0, 30	0, 2034339	0, 0290707	0, 00611288
58	4, 5, 3	0, 30	0, 0290707	0, 2034339	0, 00611288
59	5, 3, 4	0, 10	0, 2034339	0, 6103018	0, 04277730
60	5, 4, 3	0, 10	0, 6103018	0, 2034339	0, 04277730
				$\pi_{345} =$	0, 1283447
				Total	1, 0000032

From the table it can also be seen that the inclusion probabilities comply with the requirement $\pi_i = nz_i$. For instance,

$$\begin{aligned}\pi_1 &= \sum_{s_n \ni 1} P(s_n) \\ &= \pi_{123} + \pi_{124} + \pi_{125} + \pi_{134} + \pi_{135} + \pi_{145} \\ &= 0,45 = 3z_1 \quad .\end{aligned}$$

It can also be seen that the joint inclusion probabilities obtained by the example comply with (2.36). For instance,

$$\pi_{12} = \sum_{s_n \ni 1,2} P(s_n) = \pi_{123} + \pi_{124} + \pi_{125} = 0,1783235$$

and by using (2.36)

$$\pi_{12} = \frac{2}{P}\left(\sum_{i=1}^{5}\lambda_i\right)^{-2}\left(\lambda_1\lambda_2(1-z_1-z_2) + (\lambda_1 z_2 + \lambda_2 z_1)\sum_{k=3}^{5}\lambda_k\right) = 0,1783235 \quad .$$

The joint inclusion probabilities obtained are the same as those obtained by the Durbin-Sampford procedure.

Now consider condition (2.10):

(i,j)	π_{ij}	$\frac{2}{3}\pi_i\pi_j$
1,2	0,1783234	0,180
1,3	0,2733206	0,225
1,4	0,3774850	0,270
1,5	0,0708382	0,090
2,3	0,3949968	0,300
2,4	0,5166640	0,360
2,5	0,1100122	0,120
3,4	0,6591639	0,450
3,5	0,1725157	0,150
4,5	0,2466889	0,180

Although $Var_{wr}(\hat{Y}_{pps})$ exceeds $Var_{wor}(\hat{Y}_{pps})$, it can be seen that for 3 of the 10 cases relation (2.10) does not hold. Again it can be noted that this happens whenever a sample is selected which contains 2 PSUs for which the selection probabilities are low. The total probability that a sample, for which relation (2.10) does not hold, is obtained is given by

$$\pi_{123} + \pi_{124} + \pi_{125} + \pi_{135} + \pi_{145} + \pi_{235} + \pi_{245} = 0,3408391 \quad .$$

Sampford's rejective procedure is preferred to Durbin-Sampford's procedure in practice since the selection of the sample for the first is much easier than that for the latter.

2.3 NUMERICAL INVESTIGATIONS

2.3.1 Introduction

Since the expected value of the biased estimator $\hat{Var}_{wr}(\hat{Y}_{pps})$ under sampling without replacement always leads to an overestimation of $Var_{wor}(\hat{Y}_{pps})$, various artificial populations have been created to investigate under which circumstances this overestimation is the lowest. Also, since $\hat{Var}_{wr}(\hat{Y}_{pps})$ instead of the unbiased estimator $\hat{Var}_{wor}(\hat{Y}_{pps})$ is used, a comparison between these two estimators will be made.

In HSRC's practice, the number of PSUs formed in a stratum varies generally between 8 and 15, from which a sample of size 2 or 3 is normally selected. Thus for the artificial populations, N was taken as 8, 10, 12 and 15 and n as 2 and 3.

In Chapter 1 it was shown that sampling with probability proportional to some measure of size is advisable, provided these measures of size are correlated with the y_i's, since then $Var_{wor}(\hat{Y}_{pps})$ is small and if the overestimation is small, $E_{wor}\left[\hat{Var}_{wr}(\hat{Y}_{pps})\right]$ will also be small. The initial selection probabilities were thus taken as $z_i = M_i/M_0$, such that $\pi_i = nM_i/M_0$.

Since the factor $(\pi_i\pi_j - \pi_{ij})/\pi_{ij}$ in the formula of the unbiased estimator $\hat{Var}_{wor}(\hat{Y}_{pps})$ varies from sample to sample if sampling with probability proportional to MOS is applied, it introduces an instability in the variance estimation (Durbin (1967)).

For the biased estimator $\hat{Var}_{wr}(\hat{Y}_{pps})$ the coefficient of $(y_i/\pi_i - y_j/\pi_j)^2$ is constant. According to Durbin (1967), $\hat{Var}_{wr}(\hat{Y}_{pps})$ is thus preferable to $\hat{Var}_{wor}(\hat{Y}_{pps})$ in this respect.

From the examples given in Sections 2.2.2 and 2.2.3 it can be checked that the factor $(\pi_i\pi_j - \pi_{ij})/\pi_{ij}$ is large if a sample of PSUs having low selection probabilities and thus low inclusion probabilities is selected and vice versa. For instance, for the example previously used to illustrate Brewer's and Durbin's method, the following values of this factor are obtained for each possible sample:

Sample $s = (i,j)$	π_i	π_j	π_{ij}	$(\pi_i\pi_j - \pi_{ij})/\pi_{ij}$
1,2	0,4	0,8	0,2535	0,2623
1,3	0,4	0,6	0,1188	1,0202
1,4	0,4	0,2	0,0277	1,8881
2,3	0,8	0,6	0,4277	0,1223
2,4	0,8	0,2	0,1189	0,3457
3,4	0,6	0,2	0,0534	1,2472

The total probability that a sample is selected for which this factor is greater than one is in this case given by 0,1999. Thus to dampen the instability in the variance estimation when $\hat{Var}_{wor}(\hat{Y}_{pps})$ is used, some restrictions should be placed on the values of z_i. Since $z_i = M_i/M_0$, this implies that some restrictions should be placed on the measures of size. In practice it is often customary to form the PSUs in such a way that the measures of size vary only within a factor of 2, where the measure of size is taken as the number of people or households for example. This means that the N PSUs should be formed in such a way that

$$\frac{2}{3}.\frac{M_0}{N} \leq M_i \leq \frac{4}{3}.\frac{M_0}{N} \quad . \tag{2.37}$$

In this case

$$\frac{2}{3N} \leq z_i \leq \frac{4}{3N} \quad . \tag{2.38}$$

In order to test the effect of the variability of the z_i's on the bias, for each population of N PSUs the following three cases were created:

(i) for half of the N PSUs, z_i was taken as the lower bound of (2.38) and for the other half, z_i was taken as the upper bound of (2.38);

(ii) the z_i's of the N PSUs were spread evenly over the range $\left(\frac{2}{3N}; \frac{4}{3N}\right)$; and

(iii) the z_i's of all N PSUs were the same.

Denote these three cases by Case I, Case II and Case III respectively.

It is known that if some measure of size is correlated with y_i and sampling with probabilities proportional to measure of size is applied, then $Var_{wor}(\hat{Y}_{pps})$ and $Var_{wr}(\hat{Y}_{pps})$ are small. However it may be possible that sampling proportional to measure of size is applied, but the measures of size are not correlated with the y_i's. In order to test this effect, for each population created, the y_i's were chosen in such a way that

(i) y_i is approximately proportional to M_i;

(ii) y_i is approximately proportional to $1/M_i$ and

(iii) y_i is independent of M_i.

Denote these three cases by Case A, Case B and Case C respectively.

For each combination of N, Cases I, II and III and Cases A, B and C, two different populations were created. Denote these two sets of populations by Set 1 and Set 2. Thus in total 18 different populations for each N were considered. These populations are given in Appendix A. Note that in the examples only single-stage sampling has been considered since the overestimation decreases for multistage sampling. In Section 2.3.2 the degree of

overestimation for all the different populations will be investigated. In Section 2.3.3 the two estimators $\hat{Var}_{wr}(\hat{Y}_{pps})$ and $\hat{Var}_{wor}(\hat{Y}_{pps})$ will be compared.

2.3.2 An investigation into the degree of overestimation

Recall again that $\hat{Var}_{wr}(\hat{Y}_{pps})$ as an estimate of $Var_{wor}(\hat{Y}_{pps})$ results in an expected bias of size

$$\frac{n}{n-1}\left[Var_{wr}(\hat{Y}_{pps}) - Var_{wor}(\hat{Y}_{pps})\right] \quad . \tag{2.39}$$

Note that the values of bias (2.39) cannot be compared over the different populations since the value of the difference $Var_{wr}(\hat{Y}_{pps}) - Var_{wor}(\hat{Y}_{pps})$ is influenced by the magnitude of the y_i values. Thus the ratio $Var_{wr}(\hat{Y}_{pps})/Var_{wor}(\hat{Y}_{pps})$, which will always be greater than 1, will be compared over the different populations. However for each population, the bias (2.39) can be compared for $n = 2$ and $n = 3$.

The values of $Var_{wr}(\hat{Y}_{pps})/Var_{wor}(\hat{Y}_{pps})$ obtained for each population are given in Table 2.1.

Note that the values of the ratio $Var_{wr}(\hat{Y}_{pps})/Var_{wor}(\hat{Y}_{pps})$ as given in Table 2.1 roughly equal $(N-1)/(N-n)$, which is the same as for simple random sampling. This result agrees with that given by Cochran (1977).

The following conclusions and recommendations can be drawn from Table 2.1:

(i) The reduction in variance is relatively smaller for $n = 2$ than for $n = 3$. This effect is however cancelled out in the bias (2.39) by the factor $n/(n-1)$, since the examples show that for each population there is hardly any difference between the bias obtained for $n = 2$ and that obtained for $n = 3$.

(ii) The reduction in variance decreases as N increases. In order to be able to compare the bias (2.39) over N, the standardized quantity $bias/Var_{wor}(\hat{Y}_{pps})$ has been used. The bias also decreases as N increases.

(iii) If M_i is correlated with y_i (Case A), the values of the ratio obtained for Cases II and III are approximately the same. However the values of the ratio obtained for Case I always exceed those obtained for Cases II and III. It can also be noted that if $N \geq 10$, the difference in the values of the ratio $Var_{wr}(\hat{Y}_{pps})/Var_{wor}(\hat{Y}_{pps})$ and thus the difference in the bias between Cases I, II and III is relatively small.
Thus if it is known that M_i is correlated with y_i, it is advisable that 10 or more PSUs are formed. In this case the PSUs can be formed in any manner, as long as the different M_i's lie within the limits (2.37).

TABLE 2.1

VALUES OF $Var_{wr}(\hat{Y}_{pps})/Var_{wor}(\hat{Y}_{pps})$
OBTAINED FOR EACH POPULATION

			N							
			8		10		12		15	
			n		*n*		*n*		*n*	
Case	*Set*	*Case*	2	3	2	3	2	3	2	3
I	1	*A*	1,207	1,518	1,127	1,292	1,115	1,258	1,087	1,191
		B	1,143	1,350	1,108	1,250	1,087	1,194	1,064	1,138
		C	1,141	1,340	1,107	1,244	1,085	1,189	1,063	1,136
	2	*A*	1,215	1,548	1,142	1,331	1,115	1,259	1,091	1,201
		B	1,143	1,350	1,108	1,249	1,087	1,194	1,064	1,138
		C	1,153	1,371	1,110	1,254	1,086	1,190	1,066	1,143
II	1	*A*	1,164	1,393	1,126	1,287	1,096	1,213	1,085	1,187
		B	1,155	1,374	1,100	1,225	1,100	1,223	1,070	1,152
		C	1,176	1,428	1,119	1,273	1,092	1,024	1,082	1,179
	2	*A*	1,173	1,416	1,114	1,257	1,114	1,258	1,076	1,164
		B	1,150	1,362	1,112	1,254	1,098	1,220	1,071	1,153
		C	1,136	1,320	1,121	1,275	1,088	1,193	1,072	1,155
III	1	*A*	1,167	1,400	1,125	1,286	1,100	1,222	1,077	1,167
		B	1,167	1,400	1,125	1,286	1,100	1,222	1,077	1,167
		C	1,167	1,400	1,125	1,286	1,100	1,222	1,077	1,167
	2	*A*	1,167	1,400	1,125	1,286	1,100	1,222	1,077	1,167
		B	1,167	1,400	1,125	1,286	1,100	1,222	1,077	1,167
		C	1,167	1,400	1,125	1,286	1,100	1,222	1,077	1,167

(iv) If M_i is correlated with $1/y_i$ (Case B), the values of $Var_{wr}(\hat{Y}_{pps})/Var_{wor}(\hat{Y}_{pps})$ are the largest for Case III and the smallest for Case I. Thus, the larger the difference in the values of the measures of size the smaller the bias seems to be, if the measures of size are correlated with $1/y_i$.

(v) If the measures of size are independent of the y_i's (Case C), the same result as in (iv) is obtained.

(vi) If the N PSUs are formed according to Case I, the values obtained for the ratio are much larger if M_i is correlated with y_i than if M_i is independent of y_i or M_i is correlated with $1/y_i$. The same result also holds for the bias.

(vii) If the N PSUs are formed according to Case II, the values obtained for the ratio are mostly larger if M_i is correlated with y_i than if M_i is independent of y_i or M_i is correlated with $1/y_i$.

(viii) If the N PSUs are formed according to Case III, the values of the ratio obtained are the same for Cases A, B and C.

The following recommendations can thus be made in order for the overestimation to be smaller:

(i) In the range $8 \leq N \leq 15$, N should be as large as possible.

(ii) If it is known that M_i is correlated with y_i, N should be greater or equal to 10, since in this case it does not matter much how the PSUs are formed, as long as the N M_i's lie within the limits (2.37).

(iii) If there is any doubt whether M_i is correlated with y_i, it is advisable that all PSUs are of equal measure of size. However since this is normally impracticable, it is recommended that the measures of size are spread evenly over the range $\left(\frac{2}{3}\frac{M_0}{N}, \frac{4}{3}\frac{M_0}{N}\right)$.

2.3.3 Comparison between $\hat{Var}_{wr}(\hat{Y}_{pps})$ and $\hat{Var}_{wor}(\hat{Y}_{pps})$

For each population created, $\hat{Var}_{wr}(\hat{Y}_{pps})$ and $\hat{Var}_{wor}(\hat{Y}_{pps})$ have been calculated for each possible sample of size 2 and size 3.

It has been found that $\hat{Var}_{wr}(\hat{Y}_{pps})$ always exceeded $\hat{Var}_{wor}(\hat{Y}_{pps})$ except if $N = 8$ and $n = 3$ and the PSUs were formed according to Case I. However this exception occurred only for 4 of the 56 possible samples and the total probability that one of these samples is selected is only 0,013. Since any estimator of a variance is subject to variation it is possible that $\hat{Var}_{wr}(\hat{Y}_{pps})$ over- or underestimates $Var_{wr}(\hat{Y}_{pps})$. However since in complex sampling the estimation of the total variance is the sum of these variances over many strata, this variation tends to cancel out and the use of $\hat{Var}_{wr}(\hat{Y}_{pps})$ is expected to result in an overestimation of $Var_{wor}(\hat{Y}_{pps})$.

In order to investigate under which circumstances $\hat{Var}_{wr}(\hat{Y}_{pps})$ overestimates $\hat{Var}_{wor}(\hat{Y}_{pps})$ the least, the ratio

$$\hat{Var}_{wr}(\hat{Y}_{pps})/\hat{Var}_{wor}(\hat{Y}_{pps}) \tag{2.40}$$

has been calculated for each possible sample.

The minimum values of the ratio for each population is given in Table 2.2 and the maximum values in Table 2.3. The mean values and standard deviations of the ratio for each population are listed in Tables 2.4 and 2.5 respectively.

From inspection of these Tables it is clear that the values of $\hat{Var}_{wr}(\hat{Y}_{pps})/\hat{Var}_{wor}(\hat{Y}_{pps})$ are greatly influenced by n and N and only to a lesser extend by the Cases I, II and II and Cases A, B and C. It can be noted that as N approaches 15, the smaller the values of the maximum, the mean and the standard deviation of the ratio become. The values of these properties of the ratio are also smaller for $n = 2$ than for $n = 3$. The minimum value of the ratio is very close to one except if $N = 8$ and $n = 3$ and if the PSUs formed are of equal measure of size. Also, the maximum value of the ratio is the largest if $N = 8$ and $n = 3$. It even exceeds 2 if the PSUs are not of equal measure of size.

In order to achieve a smaller overestimation of $Var_{wr}(\hat{Y}_{pps})$ when the biased estimator $\hat{Var}_{wr}(\hat{Y}_{pps})$ is used, it is thus advisable that $10 \leq N \leq 15$ and $n = 2$.

BEST COPY AVAILABLE

TABLE 2.2

MINIMUM VALUES OF $\hat{Var}_{wr}(\hat{Y}_{pps})/\hat{Var}_{wor}(\hat{Y}_{pps})$
FOR EACH POPULATION

			N							
			8		10		12		15	
			n		*n*		*n*		*n*	
Case	*Set*	*Case*	2	3	2	3	2	3	2	3
I	1	*A*	1,000	0,962	1,012	1,000	1,000	1,014	1,000	1,015
		B	1,000	0,962	1,012	1,000	1,000	1,014	1,013	1,015
		C	1,000	0,962	1,012	1,000	1,000	1,014	1,013	1,015
	2	*A*	1,000	0,962	1,012	1,000	1,016	1,014	1,000	1,015
		B	1,000	0,962	1,012	1,000	1,013	1,014	1,000	1,015
		C	1,000	0,962	1,012	1,000	1,016	1,014	1,014	1,015
II	1	*A*	1,000	1,143	1,043	1,066	1,000	1,134	1,000	1,000
		B	1,100	1,240	1,044	1,064	1,079	1,122	1,000	1,096
		C	1,101	1,195	1,043	1,071	1,000	1,245	1,000	1,082
	2	*A*	1,051	1,092	1,084	1,156	1,059	1,110	1,060	1,098
		B	1,051	1,147	1,084	1,181	1,059	1,106	1,060	1,100
		C	1,051	1,096	1,084	1,159	1,059	1,101	1,060	1,101
III	1	*A*	1,333	1,597	1,250	1,427	1,000	1,333	1,000	1,250
		B	1,000	1,556	1,000	1,000	1,000	1,000	1,000	1,000
		C	1,333	1,600	1,250	1,428	1,000	1,333	1,000	1,250
	2	*A*	1,333	1,600	1,250	1,429	1,200	1,333	1,153	1,250
		B	1,333	1,598	1,000	1,375	1,200	1,333	1,000	1,250
		C	1,333	1,600	1,250	1,429	1,200	1,333	1,154	1,250

TABLE 2.3

MAXIMUM VALUES OF $\hat{Var}_{wr}(\hat{Y}_{pps})/\hat{Var}_{wor}(\hat{Y}_{pps})$
FOR EACH POPULATION

			N							
			8		10		12		15	
			n		*n*		*n*		*n*	
Case	*Set*	*Case*	2	3	2	3	2	3	2	3
I	1	*A*	1,670	2,462	1,474	1,903	1,359	1,651	1,269	1,437
		B	1,674	2,476	1,477	1,913	1,358	1,800	1,269	1,438
		C	1,667	2,462	1,466	1,900	1,358	1,650	1,255	1,436
	2	*A*	1,667	2,462	1,466	1,900	1,358	1,650	1,255	1,436
		B	1,670	2,461	1,467	1,901	1,358	1,662	1,269	1,438
		C	1,667	2,462	1,466	1,900	1,359	1,650	1,255	1,436
II	1	*A*	1,631	2,110	1,445	1,806	1,367	1,637	1,270	1,447
		B	1,632	2,231	1,444	1,774	1,367	1,596	1,269	1,420
		C	1,631	2,118	1,445	1,805	1,367	1,612	1,270	1,445
	2	*A*	1,691	2,362	1,440	1,714	1,380	1,647	1,251	1,406
		B	1,691	2,304	1,440	1,737	1,380	1,664	1,252	1,394
		C	1,691	2,337	1,440	1,780	1,380	1,672	1,251	1,416
III	1	*A*	1,333	1,604	1,250	1,444	1,200	1,333	1,154	1,250
		B	1,333	1,750	1,250	1,444	1,200	1,333	1,154	1,250
		C	1,333	1,607	1,250	1,444	1,200	1,333	1,154	1,250
	2	*A*	1,333	1,600	1,250	1,429	1,200	1,333	1,154	1,250
		B	1,333	1,605	1,250	1,444	1,200	1,333	1,154	1,250
		C	1,333	1,600	1,250	1,429	1,200	1,333	1,154	1,250

TABLE 2.4

MEAN VALUES OF $\hat{Var}_{wr}(\hat{Y}_{pps})/\hat{Var}_{wor}(\hat{Y}_{pps})$
FOR EACH POPULATION

			N							
			8		10		12		15	
			n		n		n		n	
Case	*Set*	*Case*	2	3	2	3	2	3	2	3
I	1	*A*	1,306	1.559	1,225	1,374	1,175	1,300	1,135	1,225
		B	1,304	1,549	1,224	1,387	1,173	1,300	1,139	1,220
		C	1,306	1,517	1,226	1,370	1,175	1,285	1,139	1,213
	2	*A*	1,306	1,578	1,227	1,384	1,180	1,298	1,136	1,228
		B	1,305	1,550	1,226	1,387	1,180	1,298	1,136	1,220
		C	1,306	1,535	1,226	1,375	1,180	1,287	1,139	1,215
II	1	*A*	1,311	1,574	1,243	1,401	1,192	1,316	1,144	1,246
		B	1,324	1,578	1,240	1,388	1,194	1,324	1,144	1,237
		C	1,333	1,602	1,240	1,405	1,189	1,318	1,148	1,247
	2	*A*	1,324	1,548	1,244	1,402	1,192	1,329	1,150	1,241
		B	1,319	1,556	1,244	1,408	1,192	1,320	1,150	1,239
		C	1,320	1,536	1,244	1,410	1,191	1,302	1,150	1,237
III	1	*A*	1,333	1,600	1,250	1,429	1,197	1,333	1,152	1,250
		B	1,310	1,609	1,228	1,424	1,185	1,332	1,135	1,247
		C	1,333	1,600	1,250	1,429	1,194	1,333	1,151	1,250
	2	*A*	1,333	1,600	1,250	1,429	1,200	1,333	1,154	1,250
		B	1,333	1,600	1,244	1,428	1,200	1,333	1,152	1,250
		C	1,333	1,600	1,250	1,429	1,200	1,333	1,154	1,250

TABLE 2.5

STANDARD DEVIATIONS OF $\hat{Var}_{wr}(\hat{Y}_{pps})/\hat{Var}_{wor}(\hat{Y}_{pps})$
FOR EACH POPULATION

			N							
			8		10		12		15	
			n		*n*		*n*		*n*	
Case	*Set*	*Case*	2	3	2	3	2	3	2	3
I	1	*A*	0,224	0,420	0,154	0,254	0,120	0,184	0,086	0,124
		B	0,220	0,291	0,150	0,187	0,113	0,140	0,084	0,095
		C	0,224	0,323	0,153	0,202	0,116	0,144	0,083	0,105
	2	*A*	0,224	0,411	0,153	0,256	0,116	0,182	0,085	0,123
		B	0,222	0,293	0,153	0,187	0,116	0,137	0,083	0,096
		C	0,224	0,343	0,153	0,205	0,116	0,145	0,083	0,105
II	1	*A*	0,146	0,224	0,103	0,177	0,071	0,101	0,053	0,071
		B	0,137	0,212	0,101	0,145	0,068	0,092	0,053	0,064
		C	0,131	0,237	0,102	0,142	0,070	0,096	0,050	0,065
	2	*A*	0,170	0,287	0,081	0,113	0,076	0,120	0,043	0,064
		B	0,165	0,224	0,081	0,125	0,076	0,101	0,043	0,056
		C	0,165	0,232	0,081	0,131	0,076	0,103	0,043	0,063
III	1	*A*	0,000	0,001	0,000	0,001	0,025	0,000	0,015	0,000
		B	0,087	0,035	0,072	0,042	0,053	0,023	0,049	0,026
		C	0,000	0,001	0,000	0,001	0,035	0,000	0,021	0,000
	2	*A*	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
		B	0,000	0,001	0,037	0,005	0,000	0,000	0,015	0,000
		C	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000

A sample size of 2 is recommended especially if $N = 8$.

From inspection of Table 2.4 it can be seen that the mean value of $\hat{Var}_{wr}(\hat{Y}_{pps})/\hat{Var}_{wor}(\hat{Y}_{pps})$ is only influenced by n and N, since for each combination of n and N, the mean is approximately the same for each population.

Furthermore, from Tables 2.2, 2.3 and 2.5 it can be noted that the values of y_i have hardly any influence on the maximum value, minimum value and standard deviation of the ratio, especially if $n = 2$. Therefore, irrespective of whether M_i is correlated with y_i or not, sampling with probability proportional to measure of size can be applied.

Now consider Cases I, II and III. The minimum value of the ratio is the smallest for Case I and the largest for Case III, while the maximum value of the ratio is the smallest for Case III and the largest for Case I. The range of the values of the ratio is thus the smallest for Case III and the largest for Case I. Also, for Case III, the standard deviation of the ratio equals zero or is very close to zero. This implies that for Case III, each $\hat{Var}_{wr}(\hat{Y}_{pps})$ exceeds each $\hat{Var}_{wor}(\hat{Y}_{pps})$ to the same degree for each combination of n and N, irrespective of y_i. The standard deviation is the largest for Case I.

Therefore, if it is only possible to form a small number of PSUs in a stratum or if n exceeds 2, it is advisable that all the PSUs are of equal measure of size, i.e. Case III. However, since this is seldom possible in practice, it is recommended that the measures of size of the N PSUs are evenly spread over the range $\left(\frac{2}{3}\frac{M_0}{N}; \frac{4}{3}\frac{M_0}{N}\right)$, i.e Case II. Thus, the smaller the variation in M_i, the smaller the overestimation will be.

CHAPTER 3

SAMPLING WITHOUT REPLACEMENT IN THE CASE WHERE $\pi_i \neq nz_i$

3.1 THE EFFECT OF AN INAPPROPRIATE VARIANCE ESTIMATOR

If sampling without replacement is applied, an unbiased estimator of the population total is given by $\hat{Y}_{HT}$. If however in practice $\hat{Y}_{pps}$ instead of $\hat{Y}_{HT}$ is used and the sampling scheme applied does not ensure that $\pi_i = nz_i$, $\hat{Y}_{pps}$ is not identical to $\hat{Y}_{HT}$ and the population total is estimated with a bias. In this case the mean square error instead of the variance of $\hat{Y}_{pps}$ should be estimated. Denote the mean square error of $\hat{Y}_{pps}$ under sampling without replacement by $MSE_{wor}(\hat{Y}_{pps})$. In practice, the biased estimator $\hat{Var}_{wr}(\hat{Y}_{pps})$ is used to estimate $MSE_{wor}(\hat{Y}_{pps})$.

Now, in order for the Kish assertion to be true, it seems reasonable that the relation

$$E_{wor}\left[\hat{Var}_{wr}(\hat{Y}_{pps})\right] > MSE_{wor}(\hat{Y}_{pps}) = Var_{wor}(\hat{Y}_{pps}) + bias^2 \tag{3.1}$$

should hold. This means, the use of $\hat{Var}_{wr}(\hat{Y}_{pps})$ will on the average result in a conservative estimate of $MSE_{wor}(\hat{Y}_{pps})$.

Since the mean square error can be written as

$$\begin{aligned} MSE_{wor}(\hat{Y}_{pps}) &= E_{wor}(\hat{Y}_{pps} - Y)^2 \\ &= E_{wor}\left[\hat{Y}_{pps} - E_{wor}(\hat{Y}_{pps}) + E_{wor}(\hat{Y}_{pps}) - Y\right]^2 \\ &= Var_{wor}(\hat{Y}_{pps}) + \left[E_{wor}(\hat{Y}_{pps}) - Y\right]^2 \quad , \end{aligned}$$

the bias in (3.1) is given by

$$bias = E_{wor}(\hat{Y}_{pps}) - Y \tag{3.2a}$$

$$= E_{wor}(\hat{Y}_{pps} - \hat{Y}_{HT}) \quad . \tag{3.2b}$$

Now consider the bias as given in (3.2b). Yates and Grundy (1953) have shown that in the case of sampling without replacement the use of $\hat{Y}_{pps}$ instead of $\hat{Y}_{HT}$ results in a very small

bias. They state that the bias arises in part from the correlation between z_i and y_i/z_i, where a positive correlation will give a negative bias and vice versa. They state further that the bias will tend to cancel out when several strata are used. From the investigations done in Section 3.3 it has also been found that the bias is very small.

Since the bias is negligible, relation (3.1) can thus be replaced by

$$E_{wor}\left[\hat{Var}_{wr}(\hat{Y}_{pps})\right] > Var_{wor}(\hat{Y}_{pps}) \quad . \tag{3.3}$$

Thus, if relation (3.3) holds, the use of $\hat{Var}_{wr}(\hat{Y}_{pps})$ will lead on the average to an overestimation of $Var_{wor}(\hat{Y}_{pps})$.

Relation (3.3) will be tested numerically in Section 3.3 for two sampling procedures discussed in Section 3.2.

3.2 SAMPLING PROCEDURES FOR WHICH $\pi_i \neq nz_i$

Two different sampling procedures are considered. Note that for both procedures a sample of n distinct PSUs is obtained but the inclusion probability of PSU i is not equal to nz_i. The first procedure considered is called successive sampling and it will be shown that this method is identical to that of sampling without replacement, while the second is a rejective sampling procedure.

3.2.1 Successive Sampling

Successive sampling is sampling with replacement where at each draw one PSU is selected with selection probability z_i, but if at any draw a PSU already selected is drawn again, this PSU is ignored and another PSU is selected. This process is continued until a sample of n distinct PSUs is obtained.

Define a *draw* as a series of *attempts* necessary to obtain a PSU not already selected at a previous draw. Thus to obtain a sample of n distinct PSUs, n draws are necessary, but for each draw an unknown number of attempts are needed. The drawing probabilities for each draw do not equal z_i anymore.

The first PSU is still selected with drawing probability

$$p_{i_1} = z_i \quad . \tag{3.4}$$

The drawing probability of PSU $j \neq i$ in the second draw when PSU i has been selected in the first draw, is given by

$$\begin{aligned} p_{j_2} =& \text{P(select PSU } j \text{ at the first attempt)} \\ &+\text{P(select PSU } i \text{ at the first attempt and PSU } j \text{ at the second)} \\ &+\text{P(select PSU } i \text{ at the first two attempts and PSU } j \text{ at the third)} \\ &+\ldots \\ =& z_j + z_i z_j + z_i^2 z_j + z_i^3 z_j + \cdots \\ =& z_j(1 + z_i + z_i^2 + z_i^3 + \ldots) \\ =& \frac{z_j}{1 - z_i} \quad . \end{aligned} \tag{3.5}$$

The drawing probability of PSU $k \neq i, j$ in the third draw, when PSUs i and j have been selected in the first two draws, is given by

$$\begin{aligned} p_{k_3} =& \text{P(select PSU } k \text{ at the first attempt)} \\ &+\text{P(select PSU } i \text{ at the first attempt and PSU } k \text{ at the second)} \\ &+\text{P(select PSU } j \text{ at the first attempt and PSU } k \text{ at the second)} \\ &+\text{P(select PSU } i \text{ at the first two attempts and PSU } k \text{ at the third)} \\ &+\text{P(select PSU } j \text{ at the first two attempts and PSU } k \text{ at the third)} \\ &+\text{P(select PSU } i \text{ at the first attempt, PSU } j \text{ at the second and PSU } k \text{ at the third)} \\ &+\text{P(select PSU } j \text{ at the first attempt, PSU } i \text{ at the second and PSU } k \text{ at the third)} \\ &+\ldots \\ =& z_k + z_i z_k + z_j z_k + z_i^2 z_k + z_j^2 z_k + z_i z_j z_k + z_j z_i z_k + \ldots \\ =& z_k + (z_i + z_j) z_k + (z_i^2 + z_j^2 + z_i z_j + z_j z_i) z_k + \ldots \\ =& z_k \left[1 + (z_i + z_j) + (z_i + z_j)^2 + \ldots\right] \\ =& \frac{z_k}{(1 - z_i - z_j)} \quad . \end{aligned} \tag{3.6}$$

Continuing in the same manner, the drawing probability of PSU t in the n-th draw when PSUs $i, j, k, \ldots, r$ have been selected in the previous $n-1$ draws, is given by

$$p_{t_n} = \frac{z_t}{(1 - z_i - z_j - \ldots - z_r)} \quad . \tag{3.7}$$

Now consider sampling without replacement where the drawing probabilities defined for each succeeding draw are based on the initial selection probabilities of the PSUs not already selected.

This means that for a sample of size n, the first PSU is selected with drawing probability

$$p_{i_1} = z_i \quad , \tag{3.8}$$

the second with drawing probability

$$p_{j_2} = \frac{z_j}{1 - z_i} \quad , \tag{3.9}$$

given PSU i has been selected, the third with drawing probability

$$p_{k_3} = \frac{z_k}{(1 - z_i - z_j)} \quad , \tag{3.10}$$

given PSUs i and j have been selected and the n-th with drawing probability

$$p_{t_n} = \frac{z_t}{(1 - z_i - z_j - \ldots - z_r)} \quad , \tag{3.11}$$

where PSUs $i, j, k, \ldots, r, t$ are drawn in this order.

It can thus be seen that the drawing probabilities for successive sampling, defined by (3.4) to (3.7), are identical to (3.8) to (3.10), respectively, for the usual method of sampling without replacement . This means that the theory of successive sampling is identical to that of sampling without replacement . However the method of selecting the sample in practice differs for the two procedures. Note that for sampling without replacement , only one attempt is needed for each draw, while for successive sampling more attempts may be necessary.

Now, the probability of selecting a sample containing n distinct PSUs is given by

$$P(s_n) = \frac{z_i z_j z_k \ldots z_t}{(1 - z_i)(1 - z_i - z_j) \ldots (1 - z_i - z_j - \ldots - z_r)} \quad , \tag{3.12}$$

where PSUs $i, j, k, \ldots, r, t$ are drawn in this order.

If the order of draw is ignored, the probability of selecting a sample containing n distinct PSUs is given by

$$P(s) = \left(\prod_{i \in s} z_i \right) \sum_{(1,2,\cdots,r,t)}^{*} [(1 - z_i)(1 - z_i - z_j) \ldots (1 - z_i - z_j - \ldots - z_r)]^{-1} \quad , \tag{3.13}$$

where the summation extends over all $n!$ permutations of members of s.

Note that the second example of Section 1.3 can be considered as a successive sampling scheme where $n = 2$. The formulas for the calculation of the inclusion and joint inclusion probabilities are given in that section.

Since the formulas for the calculation of the inclusion and joint inclusion probabilities are rather complicated for the general n, they are given here only for $n = 3$. In order to illustrate successive sampling, the example of Section 2.2.3 is used.

By using (1.9), the inclusion probability of PSU i in the sample is given by

$$\begin{aligned}\pi_i =&\delta_i(1) + \delta_i(2) + \delta_i(3)\\ =&z_i + \sum_{j\neq i}^{N} z_j \frac{z_i}{(1-z_j)} + \sum_{j\neq i}^{N}\sum_{k\neq(j,i)}^{N} z_j \frac{z_k}{(1-z_j)}\frac{z_i}{(1-z_j-z_k)}\\ =&z_i\left(1 + \sum_{j\neq i}^{N}\frac{z_j}{1-z_j} + \sum_{j\neq i}^{N}\frac{z_j}{1-z_j}\sum_{k\neq(j,i)}^{N}\frac{z_k}{(1-z_j-z_k)}\right) \quad . \end{aligned} \tag{3.14}$$

It can be noted that $\pi_i \neq nz_i$.

By using (1.11), the joint inclusion probability of both PSUs i and j in the sample is given by

$$\begin{aligned}\pi_{ij} =&\delta_{ij}(1,2) + \delta_{ij}(1,3) + \delta_{ij}(2,1) + \delta_{ij}(2,3) + \delta_{ij}(3,1) + \delta_{ij}(3,2)\\ =&z_i\frac{z_j}{(1-z_i)} + z_i\sum_{k=1(\neq(i\neq j))}^{N}\frac{z_k}{(1-z_i)}\frac{z_j}{(1-z_i-z_k)}\\ &+z_j\frac{z_i}{(1-z_j)} + \sum_{k=1(\neq(i\neq j))}^{N} z_k\frac{z_i}{(1-z_k)}\frac{z_j}{(1-z_i-z_k)}\\ &+z_j\sum_{k=1(\neq(i\neq j))}^{N}\frac{z_k}{(1-z_j)}\frac{z_i}{(1-z_j-z_k)} + \sum_{k=1(\neq(i\neq j))}^{N} z_k\frac{z_j}{(1-z_k)}\frac{z_i}{(1-z_j-z_k)}\\ =&z_iz_j\left[\frac{1}{(1-z_i)} + \frac{1}{(1-z_j)} + \sum_{k=1(\neq(i\neq j))}^{N}\frac{z_k}{(1-z_k)}\left(\frac{1}{(1-z_i-z_k)} + \frac{1}{(1-z_j-z_k)}\right)\right.\\ &\left. + \sum_{k=1(\neq(i\neq j))}^{N} z_k\left(\frac{1}{(1-z_i)(1-z_i-z_k)} + \frac{1}{(1-z_j)(1-z_j-z_k)}\right)\right] \quad . \end{aligned} \tag{3.15}$$

Since successive sampling is identical to ordinary sampling without replacement , the relations defined by (1.13) also hold for this sampling scheme (see Cochran (1977)).

Example

One of the following 60 samples can be selected from the universe given on page 38, Section 2.2.3:

Sample no.	*Sample*	*Drawing*		*Probabilities*	$P(s_n)$
	$s_n = (i, j, k)$	p_{i_1}	p_{j_2}	p_{k_3}	$p_{i_1} \cdot p_{j_2} \cdot p_{k_3}$
1	1, 2, 3	0, 15	0, 2352941	0, 3846153	0, 0135746
2	1, 3, 2	0, 15	0, 2941176	0, 3333333	0, 0147058
3	2, 1, 3	0, 20	0, 1875000	0, 3846153	0, 0144230
4	2, 3, 1	0, 20	0, 3125000	0, 2727272	0, 0170454
5	3, 1, 2	0, 25	0, 2000000	0, 3333333	0, 0166666
6	3, 2, 1	0, 25	0, 2666666	0, 2727272	0, 0181818
				$\pi_{123} =$	0, 0945972
7	1, 2, 4	0, 15	0, 2352941	0, 4615384	0, 0162895
8	1, 4, 2	0, 15	0, 3529411	0, 3636363	0, 0192513
9	2, 1, 4	0, 20	0, 1875000	0, 4615384	0, 0173076
10	2, 4, 1	0, 20	0, 3750000	0, 3000000	0, 0225000
11	4, 1, 2	0, 30	0, 2142857	0, 3636363	0, 0233766
12	4, 2, 1	0, 30	0, 2857142	0, 3000000	0, 0257142
				$\pi_{124} =$	0, 1244392
13	1, 2, 5	0, 15	0, 2352941	0, 1538461	0, 0054299
14	1, 5, 2	0, 15	0, 1176470	0, 2666666	0, 0047059
15	2, 1, 5	0, 20	0, 1875000	0, 1538461	0, 0057692
16	2, 5, 1	0, 20	0, 1250000	0, 2142857	0, 0053571
17	5, 1, 2	0, 10	0, 1666666	0, 2666666	0, 0044444
18	5, 2, 1	0, 10	0, 2222222	0, 2142857	0, 0047619
				$\pi_{125} =$	0, 0304684
19	1, 3, 4	0, 15	0, 2941176	0, 5000000	0, 0220588
20	1, 4, 3	0, 15	0, 3529411	0, 4545454	0, 0240641
21	3, 1, 4	0, 25	0, 2000000	0, 5000000	0, 0250000
22	3, 4, 1	0, 25	0, 4000000	0, 3333333	0, 0333333
23	4, 1, 3	0, 30	0, 2142857	0, 4545454	0, 0292207
24	4, 3, 1	0, 30	0, 3571428	0, 3333333	0, 0357142
				$\pi_{134} =$	0, 1693911

Sample no.	Sample	Drawing		Probabilities	$P(s_n)$
	$s_n = (i, j, k)$	p_{i_1}	p_{j_2}	p_{k_3}	$p_{i_1} \cdot p_{j_2} \cdot p_{k_3}$
25	1, 3, 5	0, 15	0, 2941176	0, 1666666	0, 0073529
26	1, 5, 3	0, 15	0, 1176470	0, 3333333	0, 0058823
27	3, 1, 5	0, 25	0, 2000000	0, 1666666	0, 0083333
28	3, 5, 1	0, 25	0, 1333333	0, 2307692	0, 0076923
29	5, 1, 3	0, 10	0, 1666666	0, 3333333	0, 0055556
30	5, 3, 1	0, 10	0, 2777777	0, 2307692	0, 0064103
				$\pi_{135} =$	0, 0412267
31	1, 4, 5	0, 15	0, 3529411	0, 1818181	0, 0096257
32	1, 5, 4	0, 15	0, 1176470	0, 4000000	0, 0070588
33	4, 1, 5	0, 30	0, 2142857	0, 1818181	0, 0116883
34	4, 5, 1	0, 30	0, 1428571	0, 2500000	0, 0107142
35	5, 1, 4	0, 10	0, 1666666	0, 4000000	0, 0066667
36	5, 4, 1	0, 10	0, 3333333	0, 2500000	0, 0083333
				$\pi_{145} =$	0, 0540870
37	2, 3, 4	0, 20	0, 3125000	0, 5454545	0, 0340909
38	2, 4, 3	0, 20	0, 3750000	0, 5000000	0, 0375000
39	3, 2, 4	0, 25	0, 2666666	0, 5454545	0, 0363636
40	3, 4, 2	0, 25	0, 4000000	0, 4444444	0, 0444444
41	4, 2, 3	0, 30	0, 2857142	0, 5000000	0, 0428571
42	4, 3, 2	0, 30	0, 3571428	0, 4444444	0, 0476190
				$\pi_{234} =$	0, 2428750
43	2, 3, 5	0, 20	0, 3125000	0, 1818181	0, 0113636
44	2, 5, 3	0, 20	0, 1250000	0, 3571428	0, 0089293
45	3, 2, 5	0, 25	0, 2666666	0, 1818181	0, 0121212
46	3, 5, 2	0, 25	0, 1333333	0, 3076923	0, 0102564
47	5, 2, 3	0, 10	0, 2222222	0, 3571428	0, 0079365
48	5, 3, 2	0, 10	0, 2777777	0, 3076923	0, 0085470
				$\pi_{235} =$	0, 0591540
49	2, 4, 5	0, 20	0, 3750000	0, 2000000	0, 0150000
50	2, 5, 4	0, 20	0, 1250000	0, 4285714	0, 0107142
51	4, 2, 5	0, 30	0, 2857142	0, 2000000	0, 0171428
52	4, 5, 2	0, 30	0, 1428571	0, 3333333	0, 0142857
53	5, 2, 4	0, 10	0, 2222222	0, 4285714	0, 0095238
54	5, 4, 2	0, 10	0, 3333333	0, 3333333	0, 0111111
				$\pi_{245} =$	0, 0777776

Sample no.	*Sample*	*Drawing*		*Probabilities*	$P(s_n)$
	$s_n = (i,j,k)$	p_{i_1}	p_{j_2}	p_{k_3}	$p_{i_1} \cdot p_{j_2} \cdot p_{k_3}$
55	3, 4, 5	0, 25	0, 4000000	0, 2222222	0, 0222222
56	3, 5, 4	0, 25	0, 1333333	0, 4615384	0, 0153846
57	4, 3, 5	0, 30	0, 3571428	0, 2222222	0, 0238095
58	4, 5, 3	0, 30	0, 1428571	0, 4166666	0, 0178571
59	5, 3, 4	0, 10	0, 2777777	0, 4615384	0, 0128205
60	5, 4, 3	0, 10	0, 3333333	0, 4166666	0, 0138888
				$\pi_{345} =$	0, 1059827
				Total	0, 9999989

From the table it can be seen that the inclusion probabilities comply with (3.14). For instance,

$$\begin{aligned}\pi_1 &= \sum_{s_n \ni 1} P(s_n) \\ &= \pi_{123} + \pi_{124} + \pi_{125} + \pi_{134} + \pi_{135} + \pi_{145} \\ &= 0,5142096\end{aligned}$$

and by using (3.14)

$$\pi_1 = z_1 \left[1 + \sum_{j=2}^{5} \frac{z_j}{(1-z_j)} \left(1 + \sum_{k \neq j}^{5} \frac{z_k}{(1-z_j-z_k)} \right) \right] = 0,5142109 \quad .$$

The following inclusion probabilities are obtained:

$$\pi_1 = 0,5142 \quad , \quad \pi_2 = 0,6293 \quad , \quad \pi_3 = 0,7132 \quad , \quad \pi_4 = 0,7746 \quad , \quad \pi_5 = 0,3687 \quad .$$

Note that

$$\sum_{i=1}^{5} \pi_i = 3,0000 = n \quad .$$

It can also be seen that the joint inclusion probabilities obtained by the example comply with (3.15). For instance,

$$\pi_{12} = \sum_{s_n \ni 1,2} P(s_n) = \pi_{123} + \pi_{124} + \pi_{125} = 0,2495048$$

and by (3.15)

$$\begin{aligned}\pi_{12} =& z_1 z_2 \left[\frac{1}{(1-z_1)} + \frac{1}{(1-z_2)} + \sum_{k=3}^{5} \frac{z_k}{(1-z_k)} \left(\frac{1}{(1-z_1-z_k)} + \frac{1}{(1-z_2-z_k)} \right) \right. \\ &\left. + \sum_{k=3}^{5} z_k \left(\frac{1}{(1-z_1)(1-z_1-z_k)} + \frac{1}{(1-z_2)(1-z_2-z_k)} \right) \right] \\ =& 0,2495055 \quad .\end{aligned}$$

The following joint inclusion probabilities are obtained:

$\pi_{12} = 0,2495$, $\pi_{13} = 0,3052$, $\pi_{14} = 0,3479$, $\pi_{15} = 0,1258$, $\pi_{23} = 0,3966$,

$\pi_{24} = 0,4451$, $\pi_{25} = 0,1674$, $\pi_{34} = 0,5182$, $\pi_{35} = 0,2064$, $\pi_{45} = 0,2378$.

Note that

$$\sum_{j \neq 1}^{5} \pi_{1j} = 1,0284 = 2(0,5142) = (n-1)\pi_1$$

and $$\sum_{i=1}^{5} \sum_{j>i}^{5} \pi_{ij} = 2,9999 = \frac{n(n-1)}{2} \quad .$$

Thus the relations given by (1.13) hold.

3.2.2 Rejective sampling

According to Hajek (1981), rejective sampling may be defined as sampling with replacement modified by the condition that if a PSU already selected at some previous draw, is drawn again, the partially built up sample is rejected and a new sample is selected. This process is continued until a sample of n distinct PSUs is obtained. The drawing probability of PSU i at any draw remains z_i, i.e. equal to the initial selection probability.

Now, the probability of selecting a sample s_n containing n distinct PSUs is given by

$$P(s_n) = c^* z_i z_j z_k \dots z_t \quad , \tag{3.16}$$

where PSUs $i, j, k, \dots, t$ are drawn in this order and the constant c^* is chosen such that $\sum P(s_n) = 1$.

Since sampling with replacement is applied, there are $n!$ different orderings of $i, j, k, \dots, t$. Thus the probability of selecting a sample containing n distinct PSUs, where the order of draw is ignored, is given by

$$P(s) = c^* n! \prod_{i \in s} z_i \quad . \tag{3.17}$$

Since in ordinary sampling with replacement

$$\sum_{s} P(s) = 1 \quad ,$$

where the summation extends over all possible samples s, the constant c^* is the reciprocal of the probability that a distinct sample is obtained. This means,

$$\begin{aligned} c^* &= [P(\text{sample elements are distinct})]^{-1} \\ &= [1 - P(\text{sample elements are not distinct})]^{-1} \quad . \end{aligned} \tag{3.18}$$

Since the formulas for the calculation of the inclusion and joint inclusion probabilities are complicated and difficult to derive for the general n, Hajek (1981) only gives approximations of π_i and π_{ij} which are more simple than the exact formulas. However the exact formulas for the calculation of π_i and π_{ij} for a sample of size 2 are given, but not derived. Since in a complex sampling scheme, the sample size seldom exceeds 3, exact formulas for the calculation of π_i and π_{ij} for $n = 2$ and $n = 3$ are now derived.

The inclusion probabilities are determined in the following way:

(i) Determine π_i for ordinary sampling with replacement . Denote this probability by ${\pi_i}^I$.

(ii) Determine all the samples which contain PSU i more than once in the case of ordinary sampling with replacement .

(iii) Determine all the samples which contain any PSU more than once together with PSU i.

(iv) Calculate the probability of selecting the samples determined in (ii) and (iii) above. Denote this probability by ${\pi_i}^{II}$.

(v) The inclusion probability of PSU i in a distinct sample is now given by

$$\pi_i^* = {\pi_i}^I - {\pi_i}^{II} \quad .$$

(vi) Since the probability of obtaining a sample containing n distinct PSUs is given by $(c^*)^{-1}$, the inclusion probability of PSU i in the case of rejective sampling is given by

$$\pi_i = c^* \pi_i^* \quad .$$

Firstly consider the case where $n = 2$. In this case, the only non-distinct samples that can be selected are the N samples of the form (i, i). The probability of selecting such a sample is given by z_i^2. Thus by (3.18)

$$c^* = \left(1 - \sum_{i=1}^{N} z_i^2\right)^{-1} \quad . \tag{3.19}$$

By using (1.8), ${\pi_i}^I$ is given by

$${\pi_i}^I = 2z_i - z_i^2 \tag{3.20}$$

and since only the sample (i, i) contains PSU i more than once,

$${\pi_i}^{II} = z_i^2 \quad . \tag{3.21}$$

Therefore,

$$\begin{aligned}\pi_i^* &= \pi_i^I - \pi_i^{II} \\ &= 2z_i(1-z_i) \end{aligned} \tag{3.22}$$

and the inclusion probability of PSU i in the case of rejective sampling is given by

$$\begin{aligned}\pi_i &= c^* \pi_i^* \\ &= 2z_i(1-z_i)\left(1-\sum_{i=1}^{N} z_i^2\right)^{-1} \end{aligned} \tag{3.23}$$

Note that

$$\begin{aligned}\sum_{i=1}^{N} \pi_i &= 2\left(1-\sum_{i=1}^{N} z_i^2\right)^{-1} \sum_{i=1}^{N} z_i(1-z_i) \\ &= 2\left(1-\sum_{i=1}^{N} z_i^2\right)^{-1}\left(1-\sum_{i=1}^{N} z_i^2\right) \\ &= 2 = n \end{aligned}$$

The joint inclusion probability of both PSU i and j in the sample can be determined as follows:

Since $n = 2$, the PSUs i and j can only appear together in the distinct samples (i,j) and (j,i).

Thus

$$\pi_{ij}^* = 2z_i z_j \tag{3.24}$$

and the probability of obtaining a sample with two distinct PSUs is given by $(c^*)^{-1}$ with c^* given by (3.19).

Therefore the joint inclusion probability in the case of rejective sampling is given by

$$\begin{aligned}\pi_{ij} &= c^* \pi_{ij}^* \\ &= 2z_i z_j \left(1-\sum_{i=1}^{N} z_i^2\right)^{-1} \end{aligned} \tag{3.25}$$

Note that

$$\begin{aligned}\sum_{i=1}^{N}\sum_{j>i}^{N}\pi_{ij} &= 2\left(1-\sum_{i=1}^{N}z_i^2\right)^{-1}\left(\sum_{i=1}^{N}z_i\sum_{j>i}^{N}z_j\right)\\ &= \left(1-\sum_{i=1}^{N}z_i^2\right)^{-1}\left(\sum_{i=1}^{N}z_i\sum_{j\neq i}^{N}z_j\right)\\ &= \left(1-\sum_{i=1}^{N}z_i^2\right)^{-1}\left(\sum_{i=1}^{N}z_i\left(\sum_{j=1}^{N}z_j-z_i\right)\right)\\ &= \left(1-\sum_{i=1}^{N}z_i^2\right)^{-1}\left(\sum_{i=1}^{N}z_i-\sum_{i=1}^{N}z_i^2\right)\\ &= 1 = \frac{n(n-1)}{2} \quad .\end{aligned}$$

Now consider the case where $n = 3$. In this case, the following non-distinct samples can be selected:

(i) the N samples of the form (i,i,i);

(ii) the N samples of the form (i,i,k), where k is any of the N PSUs except i. Note that there are 3 orderings of (i,i,k) possible.

The probability of selecting a non-distinct sample is therefore given by

$$\sum_{i=1}^{N}z_i^3+3\sum_{i=1}^{N}z_i^2(1-z_i) \tag{3.26}$$

and by (3.18),

$$\begin{aligned}c^* &= \left(1-\sum_{i=1}^{N}z_i^3-3\sum_{i=1}^{N}z_i^2(1-z_i)\right)^{-1}\\ &= \left(1-3\sum_{i=1}^{N}z_i^2+2\sum_{i=1}^{N}z_i^3\right)^{-1} \quad .\end{aligned} \tag{3.27}$$

By (1.8), ${\pi_i}^I$ is given by

$${\pi_i}^I = 3z_i-3z_i^2+z_i^3 \quad . \tag{3.28}$$

The following non-distinct samples contain PSU i:

(i) the sample (i,i,i) with probability

$$z_i^3 \quad ;$$

(ii) the $N-1$ samples (i,i,k) with probability

$$3z_i^2(1-z_i) \quad ; \quad \text{and}$$

(iii) the $N-1$ samples (i,k,k) with probability

$$3z_i\sum_{j\neq i}^{N} z_j^2 \quad .$$

Therefore

$$\pi_i^{II} = z_i^3 + 3z_i^2(1-z_i) + 3z_i\sum_{j\neq i}^{N} z_j^2 \tag{3.29}$$

and π_i^* is given by

$$\begin{aligned}\pi_i^* &= \pi_i^I - \pi_i^{II} \\ &= 3z_i - 6z_i^2 + 3z_i^3 - 3z_i\sum_{j\neq i}^{N} z_j^2 \quad .\end{aligned} \tag{3.30}$$

The inclusion probability of PSU i in the case of rejective sampling is therefore given by

$$\begin{aligned}\pi_i &= c^*\pi_i^* \\ &= 3z_i\left((1-z_i)^2 - \sum_{j\neq i}^{N} z_j^2\right)\left(1 - 3\sum_{i=1}^{N} z_i^2 + 2\sum_{i=1}^{N} z_i^3\right)^{-1} \quad .\end{aligned} \tag{3.31}$$

Note that

$$\begin{aligned}\sum_{i=1}^{N}\pi_i &= 3\left(1 - 3\sum_{i=1}^{N} z_i^2 + 2\sum_{i=1}^{N} z_i^3\right)^{-1}\sum_{i=1}^{N}\left(z_i - 2z_i^2 + z_i^3 - z_i\sum_{j\neq i}^{N} z_j^2\right) \\ &= 3\left(1 - 3\sum_{i=1}^{N} z_i^2 + 2\sum_{i=1}^{N} z_i^3\right)^{-1}\left(1 - 2\sum_{i=1}^{N} z_i^2 + \sum_{i=1}^{N} z_i^3 - \sum_{i=1}^{N} z_i\sum_{j\neq i}^{N} z_j^2\right) \\ &= 3\left(1 - 3\sum_{i=1}^{N} z_i^2 + 2\sum_{i=1}^{N} z_i^3\right)^{-1}\left(1 - 2\sum_{i=1}^{N} z_i^2 + \sum_{i=1}^{N} z_i^3 - \sum_{i=1}^{N} z_i\sum_{j=1}^{N} z_j^2 + \sum_{i=1}^{N} z_i^3\right) \\ &= 3 = n \quad .\end{aligned}$$

The joint inclusion probability of both PSUs i and j in the sample can be determined as follows:

Since $n=3$, the PSUs i and j can appear in a distinct sample only together with one other PSU and for each distinct sample, 6 orderings are possible.

Thus

$$\pi_{ij}^* = 6z_iz_j(1-z_i-z_j) \tag{3.32}$$

and the probability of obtaining a sample with three distinct PSUs is given by the reciprocal of (3.27). Therefore the joint inclusion probability in the case of rejective sampling is given by

$$\begin{aligned}\pi_{ij} &= c^* \pi^*_{ij} \\ &= 6 z_i z_j (1 - z_i - z_j) \left(1 - 3\sum_{i=1}^{N} z_i^2 + 2\sum_{i=1}^{N} z_i^3\right)^{-1} \quad . \end{aligned} \tag{3.33}$$

Note that

$$\begin{aligned}\sum_{i=1}^{N}\sum_{j>i}^{N} \pi_{ij} &= 6 \left(1 - 3\sum_{i=1}^{N} z_i^2 + 2\sum_{i=1}^{N} z_i^3\right)^{-1} \sum_{i=1}^{N}\sum_{j>i}^{N} z_i z_j (1 - z_i - z_j) \\ &= 3 \left(1 - 3\sum_{i=1}^{N} z_i^2 + 2\sum_{i=1}^{N} z_i^3\right)^{-1} \sum_{i=1}^{N}\sum_{j\neq i}^{N} z_i z_j (1 - z_i - z_j) \\ &= 3 \left(1 - 3\sum_{i=1}^{N} z_i^2 + 2\sum_{i=1}^{N} z_i^3\right)^{-1} \cdot \\ &\quad \left[\sum_{i=1}^{N} z_i \left(\sum_{j=1}^{N} z_j - z_i\right) - \sum_{i=1}^{N} z_i^2 \left(\sum_{j=1}^{N} z_j - z_i\right) - \sum_{i=1}^{N} z_i \left(\sum_{j=1}^{N} z_j^2 - z_i^2\right)\right] \\ &= 3 \left(1 - 3\sum_{i=1}^{N} z_i^2 + 2\sum_{i=1}^{N} z_i^3\right)^{-1} \cdot \\ &\quad \left(\sum_{i=1}^{N} z_i - \sum_{i=1}^{N} z_i^2 - \sum_{i=1}^{N} z_i^2 + \sum_{i=1}^{N} z_i^3 - \sum_{i=1}^{N} z_i \sum_{i=1}^{N} z_i^2 + \sum_{i=1}^{N} z_i^3\right) \\ &= 3 \left(1 - 3\sum_{i=1}^{N} z_i^2 + 2\sum_{i=1}^{N} z_i^3\right)^{-1} \left(1 - 3\sum_{i=1}^{N} z_i^2 + 2\sum_{i=1}^{N} z_i^3\right) \\ &= 3 = \frac{n(n-1)}{2} \quad . \end{aligned}$$

Rejective sampling will now be illustrated for $n = 2$ and $n = 3$, where the examples of Sections 1.3 and 2.2.3 are used, respectively.

Example 1: $n = 2$

One of the following 16 samples can be selected at any trial from the universe given on page 7, Section 1.3:

Sample no.	*Sample*	$P(s_n)$
	$s_n = (i, j)$	$z_i.z_j$
1	(1, 1)	0, 04
2	1, 2	0, 08
3	1, 3	0, 06
4	1, 4	0, 02
5	2, 1	0, 08
6	(2, 2)	0, 16
7	2, 3	0, 12
8	2, 4	0, 04
9	3, 1	0, 06
10	3, 2	0, 12
11	(3, 3)	0, 09
12	3, 4	0, 03
13	4, 1	0, 02
14	4, 2	0, 04
15	4, 3	0, 03
16	(4, 4)	0, 01
	Total	1, 00

Note that samples in brackets denote the non-distinct samples.

From this example it can be seen that the probability of obtaining a sample that is not distinct is given by

$$P(1,1) + P(2,2) + P(3,3) + P(4,4) = 0,3 \quad .$$

Therefore,

$$c^* = \frac{1}{0,7}$$

and it can be noted that this complies with (3.19), since

$$c^* = \left(1 - \sum_{i=1}^{N} z_i^2\right)^{-1} = \frac{1}{0,7} \quad .$$

Since there are 2 orderings possible for each distinct sample, the probability of obtaining a distinct sample, if the order of draw is ignored, is now given by (3.17). One of the following 6 distinct samples can be selected with probabilities:

Sample no.	*Sample*	$P(s)$
	$s=(i,j)$	$2c^*.z_i.z_j$
1	1, 2	0, 2286
2	1, 3	0, 1714
3	1, 4	0, 0571
4	2, 3	0, 3429
5	2, 4	0, 1143
6	3, 4	0, 0857
	Total	1, 0000

From the first table it can be seen that

$$\pi_1{}^I = P(1,1) + 2P(1,2) + 2P(1,3) + 2P(1,4) = 0,36$$

and $$\pi_1{}^{II} = P(1,1) = 0,04 \quad ,$$

such that

$$\pi_1^* = 0,32$$

and $$\pi_1 = \frac{0,32}{0,7} = 0,4571 \quad .$$

It can be seen that this inclusion probability complies with (3.23), since

$$\pi_1 = \frac{2(0,2)(0,8)}{0,7} = 0,4571 \quad .$$

The following inclusion probabilities are obtained:

$$\pi_1 = 0,4571 \quad , \quad \pi_2 = 0,6857 \quad , \quad \pi_3 = 0,6000 \quad , \quad \pi_4 = 0,2571 \quad .$$

Note that

$$\sum_{i=1}^{4} \pi_i = 1,9999 \approx 2 = n \quad .$$

Also, from the first table it can be noted that PSUs 1 and 2 can only appear together in the two samples $(1,2)$ and $(2,1)$. Therefore

$$\pi_{12}^* = P(1,2) + P(2,1) = 0,16$$

and $$\pi_{12} = \frac{0,16}{0,7} = 0,2286 \quad .$$

It can be seen that this joint inclusion probability complies with (3.25), since

$$\pi_{12} = \frac{2(0,2)(0,4)}{0,7} = 0,2286 \quad .$$

The following joint inclusion probabilities are obtained:

$$\pi_{12} = 0,2286 \quad , \quad \pi_{13} = 0,1714 \quad , \quad \pi_{14} = 0,0571 \quad ,$$

$$\pi_{23} = 0,3429 \quad , \quad \pi_{24} = 0,1143 \quad , \quad \pi_{34} = 0,0857 \quad .$$

Note that

$$\sum_{j \neq 1}^{4} \pi_{1j} = 0,4571 = (n-1)\pi_1$$

and $$\sum_{i=1}^{4} \sum_{j>i}^{4} \pi_{ij} = 1 = \frac{n(n-1)}{2} \quad .$$

Therefore the relations given by (1.13) hold.

Example 2: $n = 3$

One of the following 125 samples can be selected at any trial from the universe given on page 38, Section 2.2.3:

Sample no.	*Sample*	$P(s_n)$	*Sample no.*	*Sample*	$P(s_n)$
	$s_n = (i,j,k)$	$z_i.z_j.z_k$		$s_n = (i,j,k)$	$z_i.z_j.z_k$
1	(1, 1, 1)	0, 003375	16	(1, 4, 1)	0, 006750
2	(1, 1, 2)	0, 004500	17	1, 4, 2	0, 009000
3	(1, 1, 3)	0, 005625	18	1, 4, 3	0, 011250
4	(1, 1, 4)	0, 006750	19	(1, 4, 4)	0, 013500
5	(1, 1, 5)	0, 002250	20	1, 4, 5	0, 004500
6	(1, 2, 1)	0, 004500	21	(1, 5, 1)	0, 002250
7	(1, 2, 2)	0, 006000	22	1, 5, 2	0, 003000
8	1, 2, 3	0, 007500	23	1, 5, 3	0, 003750
9	1, 2, 4	0, 009000	24	1, 5, 4	0, 004500
10	1, 2, 5	0, 003000	25	(1, 5, 5)	0, 001500
11	(1, 3, 1)	0, 005625	26	(2, 1, 1)	0, 004500
12	1, 3, 2	0, 007500	27	(2, 1, 2)	0, 006000
13	(1, 3, 3)	0, 009375	28	2, 1, 3	0, 007500
14	1, 3, 4	0, 011250	29	2, 1, 4	0, 009000
15	1, 3, 5	0, 003700	30	2, 1, 5	0, 003000

Sample no.	Sample	$P(s_n)$	Sample no.	Sample	$P(s_n)$
	$s_n = (i, j, k)$	$z_i.z_j.z_k$		$s_n = (i, j, k)$	$z_i.z_j.z_k$
31	(2, 2, 1)	0, 006000	71	3, 5, 1	0, 003750
32	(2, 2, 2)	0, 008000	72	3, 5, 2	0, 005000
33	(2, 2, 3)	0, 010000	73	(3, 5, 3)	0, 006250
34	(2, 2, 4)	0, 012000	74	3, 5, 4	0, 007500
35	(2, 2, 5)	0, 004000	75	(3, 5, 5)	0, 002500
36	2, 3, 1	0, 007500	76	(4, 1, 1)	0, 006750
37	(2, 3, 2)	0, 010000	77	4, 1, 2	0, 009000
38	(2, 3, 3)	0, 012500	78	4, 1, 3	0, 011250
39	2, 3, 4	0, 015000	79	(4, 1, 4)	0, 013500
40	2, 3, 5	0, 005000	80	4, 1, 5	0, 004500
41	2, 4, 1	0, 009000	81	4, 2, 1	0, 009000
42	(2, 4, 2)	0, 012000	82	(4, 2, 2)	0, 012000
43	2, 4, 3	0, 015000	83	4, 2, 3	0, 015000
44	(2, 4, 4)	0, 018000	84	(4, 2, 4)	0, 018000
45	2, 4, 5	0, 006000	85	4, 2, 5	0, 006000
46	2, 5, 1	0, 003000	86	4, 3, 1	0, 011250
47	(2, 5, 2)	0, 004000	87	4, 3, 2	0, 015000
48	2, 5, 3	0, 005000	88	(4, 3, 3)	0, 018750
49	2, 5, 4	0, 006000	89	(4, 3, 4)	0, 022500
50	(2, 5, 5)	0, 002000	90	4, 3, 5	0, 007500
51	(3, 1, 1)	0, 005625	91	(4, 4, 1)	0, 013500
52	3, 1, 2	0, 007500	92	(4, 4, 2)	0, 018000
53	(3, 1, 3)	0, 009375	93	(4, 4, 3)	0, 022500
54	3, 1, 4	0, 011250	94	(4, 4, 4)	0, 027000
55	3, 1, 5	0, 003750	95	(4, 4, 5)	0, 009000
56	3, 2, 1	0, 007500	96	4, 5, 1	0, 004500
57	(3, 2, 2)	0, 001000	97	4, 5, 2	0, 006000
58	(3, 2, 3)	0, 012500	98	4, 5, 3	0, 007500
59	3, 2, 4	0, 015000	99	(4, 5, 4)	0, 009000
60	3, 2, 5	0, 005000	100	(4, 5, 5)	0, 003000
61	(3, 3, 1)	0, 009375	101	(5, 1, 1)	0, 002250
62	(3, 3, 2)	0, 012500	102	5, 1, 2	0, 003000
63	(3, 3, 3)	0, 015625	103	5, 1, 3	0, 003750
64	(3, 3, 4)	0, 018750	104	5, 1, 4	0, 004500
65	(3, 3, 5)	0, 006250	105	(5, 1, 5)	0, 001500
66	3, 4, 1	0, 011250	106	5, 2, 1	0, 003000
67	3, 4, 2	0, 015000	107	(5, 2, 2)	0, 004000
68	(3, 4, 3)	0, 018750	108	5, 2, 3	0, 005000
69	(3, 4, 4)	0, 022500	109	5, 2, 4	0, 006000
70	3, 4, 5	0, 007500	110	(5, 2, 5)	0, 002000

Sample no.	*Sample*	$P(s_n)$
	$s_n = (i, j, k)$	$z_i.z_j.z_k$
111	5, 3, 1	0, 003750
112	5, 3, 2	0, 005000
113	(5, 3, 3)	0, 006250
114	5, 3, 4	0, 007500
115	(5, 3, 5)	0, 002500
116	5, 4, 1	0, 004500
117	5, 4, 2	0, 006000
118	5, 4, 3	0, 007500
119	(5, 4, 4)	0, 009000
120	(5, 4, 5)	0, 003000
121	(5, 5, 1)	0, 001500
122	(5, 5, 2)	0, 002000
123	(5, 5, 3)	0, 002500
124	(5, 5, 4)	0, 003000
125	(5, 5, 5)	0, 001000
	Total	1, 000000

Note that samples in brackets denote the non-distinct samples.

From this table it can be seen that the probability of selecting a sample that is not distinct, is the sum of $P(s_n)$ over all samples in brackets, which equals 0,565.
Therefore,

$$c^* = \frac{1}{0,435}$$

and it can be noted that this complies with (3.27), since

$$c^* = \frac{1}{1 - (3)(0,225) + 2(0,055)} = \frac{1}{0,435}$$

Since there are 6 orderings possible for each distinct sample, the probability of obtaining a distinct sample, if the order of draw is ignored, is now given by (3.17). One of the following 10 distinct samples can be selected with probabilities:

Sample no.	*Sample*	$P(s)$
	$s = (i, j, k)$	$6c^*.z_i.z_j.z_k$
1	1, 2, 3	0, 1034
2	1, 2, 4	0, 1241
3	1, 2, 5	0, 0414
4	1, 3, 4	0, 1552
5	1, 3, 5	0, 0517
6	1, 4, 5	0, 0621
7	2, 3, 4	0, 2069
8	2, 3, 5	0, 0690
9	2, 4, 5	0, 0828
10	3, 4, 5	0, 1034
	Total	1, 0000

From inspection of the first table it can be confirmed that:

(i) ${\pi_1}^I = 0, 385875$, which complies with (3.28), since

$$ {\pi_1}^I = 3(0,15) - 3(0,15)^2 + (0,15)^3 = 0,385875 \quad ; $$

(ii) the probability of selecting the sample $(1, 1, 1)$ is given by $0, 003375$, the probability of selecting the samples of the form $(1, 1, k)$ by $0, 057375$ and the probability of selecting the samples of the form $(1, k, k)$ by $0, 091125$. Thus ${\pi_1}^{II} = 0, 151875$, which complies with (3.29), since

$$ {\pi_1}^{II} = (0,15)^3 + 3(0,15)^2(0,85) + 3(0,15)\sum_{j=2}^{5} z_j^2 = 0,151875 \quad ; $$

(iii) $\pi_1^* = 0, 234$, which complies with (3.30), since

$$ \pi_1^* = 3(0,15) - 6(0,15)^2 + 3(0,15)^3 - 3(0,15)\sum_{j=2}^{5} z_j^2 = 0,234 \quad . $$

Thus the inclusion probability of PSU 1 is given by

$$ \pi_1 = \frac{0,234}{0,435} = 0,5379 \quad . $$

The following inclusion probabilities are obtained:

$$ \pi_1 = 0,5379 \quad , \quad \pi_2 = 0,6276 \quad , \quad \pi_3 = 0,6897 \quad , \quad \pi_4 = 0,7345 \quad , \quad \pi_5 = 0,4103 \quad . $$

Note that

$$ \sum_{i=1}^{5} \pi_i = 3 = n \quad . $$

Also, from the first table it can be seen that PSUs 1 and 2 can appear together in the distinct samples of the form $(1, 2, k)$, $k \neq 1, 2$, where each of these distinct samples can be ordered in 6 different ways. Therefore,

$$\pi_{12}^{*} = 6[P(1,2,3) + P(1,2,4) + P(1,2,5)] = 0,117$$

and it can be seen that this complies with (3.32), since

$$\pi_{12}^{*} = 6(0,15)(0,2)(0,65) = 0,117 \quad .$$

Therefore, the joint inclusion probability of both PSUs 1 and 2 in a distinct sample is given by

$$\pi_{12} = \frac{0,117}{0,435} = 0,2690 \quad .$$

The following joint inclusion probabilities are obtained:

$\pi_{12} = 0,2690$, $\pi_{13} = 0,3103$, $\pi_{14} = 0,3414$, $\pi_{15} = 0,1552$, $\pi_{23} = 0,3793$,

$\pi_{24} = 0,4138$, $\pi_{25} = 0,1931$, $\pi_{34} = 0,4655$, $\pi_{35} = 0,2241$, $\pi_{45} = 0,2483$.

Note that

$$\sum_{j \neq 1}^{5} \pi_{1j} = 1,0759 = 2(0,5379) = (n-1)\pi_1 \quad ,$$

and $$\sum_{i=1}^{5} \sum_{j>i}^{5} \pi_{ij} = 3 = \frac{n(n-1)}{2} \quad .$$

Thus the relations given by (1.13) hold.

3.3 NUMERICAL INVESTIGATIONS

The populations created for the numerical investigations of Chapter 2 are also used in this section. Each investigation is done seperately for successive and rejective sampling and a comparison between these two sampling methods will be made.

In Section 3.3.1 the resulting bias when $\hat{Y}_{pps}$ instead of $\hat{Y}_{HT}$ is used under sampling without replacement , will be discussed.

To investigate whether relation (3.3) holds for successive and rejective sampling, the ratio

$$\frac{E_{wor}[\hat{Var}_{wr}(\hat{Y}_{pps})]}{Var_{wor}(\hat{Y}_{pps})} \tag{3.34}$$

is calculated for each population. If this ratio exceeds one, the Kish assertion is true on the average. The numerator of (3.3) has been calculated by using the expression

$$E_{wor}[\hat{Var}_{wr}(\hat{Y}_{pps})] = \sum_{s} \left[\hat{Var}_{wr}(\hat{Y}_{pps})\right] P(s) \tag{3.35}$$

and the denominator by using the expression

$$Var_{wor}(\hat{Y}_{pps}) = \sum_{s} \left[\hat{Y}_{pps} - \sum_{s} \hat{Y}_{pps} P(s)\right]^2 P(s) \quad , \tag{3.36}$$

where the summation extends over all possible distinct samples that can be selected from a population and where $P(s)$ is the probability of selecting such a sample. $P(s)$ is given by π_{ij} when $n = 2$ and by π_{ijk} when $n = 3$.
This investigation is considered in Section 3.3.2.

In order to investigate whether the use of $\hat{Var}_{wr}(\hat{Y}_{pps})$ under sampling without replacement overestimates the true variance of $\hat{Y}_{HT}$, the ratio

$$\frac{E_{wor}[\hat{Var}_{wr}(\hat{Y}_{pps})]}{Var_{wor}(\hat{Y}_{HT})} \tag{3.37}$$

is calculated for each population in Section 3.3.3.

In order to be able to compare the individual estimators, the ratio

$$\frac{\hat{Var}_{wr}(\hat{Y}_{pps})}{\hat{Var}_{wor}(\hat{Y}_{pps})} \tag{3.38}$$

has to be calculated. However no best unbiased estimator of $Var_{wor}(\hat{Y}_{pps})$ is readily available. Since the squared term of expression (3.36) is an unbiased, but not necessarily the best, estimator of $Var_{wor}(\hat{Y}_{pps})$, $\hat{Var}_{wor}(\hat{Y}_{pps})$ is calculated by using the expression

$$\hat{Var}_{wor}(\hat{Y}_{pps}) = \left[\hat{Y}_{pps} - \sum_{s} \hat{Y}_{pps} P(s)\right]^2 \quad . \tag{3.39}$$

This investigation is done in Section 3.3.4.

3.3.1 An investigation into the bias

Yates and Grundy (1953) have shown by means of an example that the inclusion probabilities π_i for successive sampling lie closer to nz_i than those for rejective sampling. However, they only considered a sample of size 2. In this section, samples of size 3 are also considered and it will be investigated whether the bias is also negligible in this case.

In order to investigate the bias when $\hat{Y}_{pps}$ instead of $\hat{Y}_{HT}$ is used under sampling without replacement , the ratio

$$\frac{\hat{Y}_{pps}}{\hat{Y}_{HT}} \tag{3.38}$$

is calculated for each possible sample that can be selected from a given population. For each population, the mean value and the standard deviation of this ratio is determined. These quantities are given in Tables 3.1 and 3.2 respectively.

TABLE 3.1

MEAN VALUES OF $\hat{Y}_{pps}/\hat{Y}_{HT}$
FOR EACH POPULATION

				SUCCESSIVE SAMPLING				*REJECTIVE SAMPLING*			
				N				*N*			
Case	*Set*	*Case*	*n*	8	10	12	15	8	10	12	15
I	1	*A*	2	1,009	1,007	1,005	1,004	1,015	1,012	1,010	1,007
			3	1,017	1,014	1,010	1,008	1,026	1,022	1,018	1,014
		B	2	1,018	1,014	1,011	1,008	1,032	1,025	1,020	1,015
			3	1,046	1,034	1,027	1,020	1,075	1,058	1,047	1,036
		C	2	1,014	1,011	1,008	1,006	1,025	1,019	1,015	1,011
			3	1,034	1,025	1,018	1,014	1,055	1,043	1,031	1,025
	2	*A*	2	1,008	1,006	1,005	1,004	1,014	1,011	1,009	1,007
			3	1,015	1,012	1,010	1,008	1,022	1,020	1,016	1,014
		B	2	1,018	1,014	1,011	1,008	1,032	1,025	1,020	1,015
			3	1,046	1,034	1,027	1,019	1,075	1,058	1,047	1,035
		C	2	1,014	1,011	1,008	1,006	1,024	1,020	1,015	1,011
			3	1,033	1,026	1,019	1,014	1,053	1,045	1,033	1,025

TABLE 3.1 (CONT.)

				SUCCESSIVE SAMPLING				*REJECTIVE SAMPLING*			
				N				*N*			
Case	*Set*	*Case*	*n*	8	10	12	15	8	10	12	15
II	1	*A*	2	1,003	1,003	1,002	1,001	1,005	1,005	1,003	1,002
			3	1,006	1,006	1,003	1,002	1,009	1,009	1,005	1,004
		B	2	1,006	1,006	1,003	1,003	1,011	1,011	1,006	1,005
			3	1,016	0,015	1,008	1,006	1,025	1,027	1,014	1,011
		C	2	1,005	1,005	1,003	1,002	1,008	1,010	1,005	1,004
			3	1,010	1,013	1,007	1,004	1,016	1,022	1,012	1,008
	2	*A*	2	1,004	1,002	1,002	1,001	1,008	1,004	1,004	1,002
			3	1,009	1,005	1,005	1,002	1,013	1,008	1,008	1,003
		B	2	1,009	1,004	1,004	1,002	1,016	1,007	1,008	1,004
			3	1,023	1,009	1,010	1,005	1,037	1,015	1,018	1,009
		C	2	1,007	1,003	1,003	1,001	1,012	1,004	1,005	1,002
			3	1,016	1,006	1,006	1,003	1,026	1,009	1,011	1,005
III	1	*A*	2	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000
			3	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000
		B	2	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000
			3	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000
		C	2	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000
			3	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000
	2	*A*	2	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000
			3	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000
		B	2	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000
			3	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000
		C	2	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000
			3	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000

TABLE 3.2

STANDARD DEVIATIONS OF $\hat{Y}_{PPS}/\hat{Y}_{HT}$
FOR EACH POPULATION

				SUCCESSIVE SAMPLING				*REJECTIVE SAMPLING*			
				N				*N*			
Case	*Set*	*Case*	*n*	8	10	12	15	8	10	12	15
I	1	*A*	2	0,018	0,014	0,011	0,009	0,032	0,025	0,021	0,016
			3	0,029	0,022	0,018	0,014	0,049	0,039	0,033	0,025
		B	2	0,020	0,015	0,012	0,010	0,035	0,028	0,023	0,018
			3	0,028	0,022	0,018	0,015	0,048	0,039	0,032	0,027
		C	2	0,019	0,015	0,012	0,009	0,034	0,026	0,021	0,017
			3	0,029	0,023	0,018	0,014	0,050	0,040	0,033	0,027
	2	*A*	2	0,018	0,014	0,011	0,009	0,033	0,026	0,021	0,016
			3	0,029	0,023	0,018	0,014	0,050	0,040	0,033	0,026
		B	2	0,020	0,015	0,012	0,010	0,035	0,028	0,023	0,018
			3	0,028	0,022	0,018	0,015	0,048	0,039	0,032	0,027
		C	2	0,019	0,015	0,012	0,009	0,034	0,026	0,022	0,017
			3	0,030	0,023	0,018	0,015	0,051	0,040	0,033	0,027
II	1	*A*	2	0,011	0,009	0,006	0,005	0,019	0,016	0,012	0,009
			3	0,017	0,014	0,010	0,008	0,028	0,025	0,018	0,014
		B	2	0,011	0,010	0,006	0,005	0,019	0,019	0,011	0,009
			3	0,016	0,017	0,010	0,008	0,028	0,030	0,017	0,015
		C	2	0,010	0,010	0,006	0,005	0,018	0,018	0,012	0,009
			3	0,016	0,016	0,010	0,007	0,027	0,027	0,018	0,013
	2	*A*	2	0,013	0,007	0,007	0,004	0,023	0,013	0,013	0,008
			3	0,021	0,012	0,011	0,007	0,035	0,020	0,020	0,013
		B	2	0,013	0,007	0,007	0,005	0,024	0,014	0,013	0,008
			3	0,020	0,012	0,011	0,007	0,035	0,021	0,019	0,013
		C	2	0,014	0,007	0,008	0,004	0,024	0,013	0,014	0,008
			3	0,022	0,012	0,013	0,007	0,037	0,021	0,022	0,013

TABLE 3.2 (CONT.)

				SUCCESSIVE SAMPLING				*REJECTIVE SAMPLING*			
				N				*N*			
Case	*Set*	*Case*	*n*	8	10	12	15	8	10	12	15
III	1	*A*	2	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
			3	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
		B	2	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
			3	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
		C	2	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
			3	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
	2	*A*	2	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
			3	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
		B	2	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
			3	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
		C	2	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
			3	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000

From inspection of these two tables, it is clear that the bias is indeed negligible since the mean values of the ratio $\hat{Y}_{ppz}/\hat{Y}_{HT}$ lie very close to unity and the standard deviations very close to zero. In fact, the mean values and standard deviations are equal to one and zero respectively, if all the N PSUs are of equal measure of size, i.e. Case III. This however has been expected, since in this case, $\pi_i = nz_i$. If all MOS are equal, the ratio is independent of the value of y_i and either successive or rejective sampling can be applied.

If the measures of size are evenly spread over the range (2.37) (Case II), the mean values and standard deviations of the ratio are always smaller for $n = 2$ than for $n = 3$. These quantities are also smaller in the case of successive sampling than in the case of rejective sampling. The ratio varies more if rejective sampling is applied.

In the case where half of the measures of size are equal to the lower bound and half equal to the upper bound of (2.37), i.e. Case I, the mean values and the standard deviations of the ratio lie closer to one and zero, respectively, if $n = 2$ than if $n = 3$. These values are smaller for successive sampling than for rejective sampling. Thus also in this case, successive sampling is less biased than rejective sampling.

Also, the mean values get closer to one as N increases and the standard deviations approach zero as N increases.

It is thus clear that successive sampling is superior to rejective sampling with regard to the bias. The result of Yates and Grundy (1953) is thus confirmed and it is also true for $n = 3$.

In order for the bias to be a minimum if $\hat{Y}_{pps}$ instead of $\hat{Y}_{HT}$ is used under sampling without replacement , it is thus advisable that N should be larger and a sample of size 2 should be selected by means of successive sampling.

It has thus been shown by means of examples that the bias is in fact negligible. In order to show how small this bias is relative to $Var_{wor}(\hat{Y}_{pps})$ of equation (3.1), the values of the ratio $bias^2/Var_{wor}(\hat{Y}_{pps})$ are given in Appendix B and it can be seen that these values are very close to zero.

3.3.2 An investigation into the over-/underestimation of $Var_{wor}(\hat{Y}_{pps})$

The values of the ratio (3.34) are given in Table 3.3 and it can be seen that for the examples used, this ratio always exceeds unity. Thus relation (3.3) holds, which means that the use of $\hat{Var}_{wr}(\hat{Y}_{pps})$ under sampling without replacement results on the average in an overestimation of $Var_{wor}(\hat{Y}_{pps})$ for these particular populations.

In order for this overestimation to be smaller, the following conclusions and recommendations can be made:

(i) The overestimation is relatively smaller for $n = 2$ than for $n = 3$. Also, this overestimation decreases as N increases. It can be seen that $N = 8$ together with a sample size of 3 is highly undesirable, since in this case the overestimations exceeds 50%. It is recommendable that $N \geq 10$ and $n = 2$, since then the overestimation is less than 30%. It can also be noted that the degree of overestimation is greatly influenced by the values of n and N and only to a lesser extent by Cases I, II and III and Cases A, B and C as given on page 49.

(ii) If M_i is correlated with y_i (Case A), the values of the ratio obtained for Cases II and III are very close to each other, especially if $n = 2$. However, the values obtained for Case I always exceed those obtained for Cases II and III.

TABLE 3.3

VALUES OF $E_{wor}[\hat{Var}_{wr}(\hat{Y}_{pps})]/Var_{wor}(\hat{Y}_{pps})$
FOR EACH POPULATION

				SUCCESSIVE SAMPLING				*REJECTIVE SAMPLING*			
				N				*N*			
Case	*Set*	*Case*	*n*	8	10	12	15	8	10	12	15
I	1	*A*	2	1,404	1,253	1,226	1,173	1,397	1,251	1,224	1,172
			3	1,730	1,429	1,378	1,282	1,701	1,423	1,370	1,278
		B	2	1,287	1,217	1,174	1,128	1,291	1,219	1,176	1,129
			3	1,535	1,380	1,295	1,209	1,550	1,389	1,301	1,213
		C	2	1,284	1,215	1,171	1,127	1,287	1,216	1,172	1,128
			3	1,520	1,371	1,286	1,206	1,533	1,377	1,292	1,209
	2	*A*	2	1,420	1,281	1,227	1,181	1,412	1,277	1,225	1,180
			3	1,770	1,480	1,379	1,296	1,738	1,469	1,372	1,291
		B	2	1,287	1,217	1,174	1,128	1,291	1,219	1,176	1,129
			3	1,535	1,380	1,295	1,209	1,550	1,389	1,301	1,213
		C	2	1,305	1,221	1,172	1,133	1,306	1,223	1,173	1,133
			3	1,555	1,383	1,287	1,215	1,562	1,390	1,292	1,217
II	1	*A*	2	1,328	1,250	1,192	1,170	1,328	1,249	1,193	1,170
			3	1,590	1,425	1,320	1,278	1,590	1,421	1,320	1,275
		B	2	1,310	1,201	1,198	1,140	1,310	1,203	1,198	1,141
			3	1,556	1,343	1,331	1,228	1,559	1,351	1,330	1,229
		C	2	1,348	1,237	1,185	1,163	1,346	1,237	1,185	1,163
			3	1,627	1,407	1,307	1,266	1,621	1,408	1,308	1,265
	2	*A*	2	1,345	1,228	1,227	1,152	1,344	1,230	1,225	1,152
			3	1,622	1,390	1,379	1,246	1,618	1,394	1,374	1,246
		B	2	1,300	1,224	1,196	1,142	1,301	1,225	1,196	1,142
			3	1,542	1,382	1,327	1,230	1,547	1,385	1,327	1,231
		C	2	1,276	1,241	1,177	1,144	1,280	1,242	1,178	1,144
			3	1,496	1,412	1,294	1,233	1,509	1,413	1,297	1,234

TABLE 3.3 (CONT.)

				SUCCESSIVE SAMPLING				*REJECTIVE SAMPLING*			
				N				*N*			
Case	*Set*	*Case*	*n*	8	10	12	15	8	10	12	15
III	1	*A*	2	1,333	1,250	1,200	1,154	1,333	1,250	1,200	1,154
			3	1,600	1,429	1,333	1,250	1,600	1,429	1,333	1,250
		B	2	1,333	1,250	1,200	1,154	1,333	1,250	1,200	1,154
			3	1,600	1,429	1,333	1,250	1,600	1,429	1,333	1,250
		C	2	1,333	1,250	1,200	1,154	1,333	1,250	1,200	1,154
			3	1,600	1,429	1,333	1,250	1,600	1,429	1,333	1,250
	2	*A*	2	1,333	1,250	1,200	1,154	1,333	1,250	1,200	1,154
			3	1,600	1,429	1,333	1,250	1,600	1,429	1,333	1,250
		B	2	1,333	1,250	1,200	1,154	1,333	1,250	1,200	1,154
			3	1,600	1,429	1,333	1,250	1,600	1,429	1,333	1,250
		C	2	1,333	1,250	1,200	1,154	1,333	1,250	1,200	1,154
			3	1,600	1,429	1,333	1,250	1,600	1,429	1,333	1,250

(iii) If M_i is not correlated with y_i (Cases B and C), the obtained values of the ratio are always the largest for Case III and the smallest for Case I.

(iv) If the N PSUs are formed according to Cases I and II, the values obtained for the ratio are relatively larger if M_i is correlated with y_i (Case A) than if M_i is not correlated with y_i (Cases B and C).

(v) If all the PSUs are of equal MOS (Case III), the obtained values of the ratio are the same for Cases A, B and C. Therefore, if all the measures of size are equal, the values of the ratio are independent of y_i.

(vi) The difference in overestimation between successive and rejective sampling is very small and thus negligible. In fact, if all the PSUs are of equal MOS (Case III), the overestimation is the same for both sampling methods.

The following overall recommendations can thus be made to ensure a smaller overestimation when $\hat{Var}_{wr}(\hat{Y}_{pps})$ instead of $\hat{Var}_{wor}(\hat{Y}_{pps})$ is used under sampling without replacement :

(i) In the range $8 \leq N \leq 15$, N should be greater or equal to 10 and $n = 2$.

(ii) If it is known that M_i is correlated with y_i, the N PSUs should be formed according to Cases II or III. Thus, in this case it is advisable that the M_i's vary only slightly.

(iii) If it is known that M_i is not correlated with y_i, the N PSUs should be formed in such a way that the variation in the measures of size is as large as possible, i.e. Case I.

(iv) If there exists any doubt whether M_i is correlated with y_i, it is advisable that all the MOS are equal, since then the overestimation is independent of y_i. However, since this is normally impracticable, it is recommended that the measures of size are spread evenly over the range $(\frac{2}{3}\frac{M_0}{N}, \frac{4}{3}\frac{M_0}{N})$, i.e. the variation between the different M_i's should be as small as possible.

(v) Since there is hardly any difference in the results obtained for rejective sampling and those obtained for successive sampling, either method can be applied.

3.3.3 An investigation into the over-/underestimation of $Var_{wor}(\hat{Y}_{HT})$

The values of the ratio (3.37) are given in Table 3.4. If all the PSUs are of equal MOS, this ratio is identical to the one given in (3.34), since in this case $\hat{Y}_{pps} = \hat{Y}_{HT}$. Thus, the values of the ratio for Case III are given in Table 3.3.

In the case where the PSUs are not of equal MOS (Cases I and II) and M_i is correlated with y_i (Case A), it be seen that the use of $\hat{Var}_{wr}(\hat{Y}_{pps})$ under sampling without replacement need not necessarily overestimate the true variance of $\hat{Y}_{HT}$. By comparing Tables 3.3 and 3.4, it can be seen that the ratio (3.37) may be smaller than that given by (3.34). Thus in this case, $Var_{wor}(\hat{Y}_{pps})$ may be smaller than $Var_{wor}(\hat{Y}_{HT})$.
However if a sample of size 2 is selected by means of successive sampling, the use of $\hat{Var}_{wr}(\hat{Y}_{pps})$ under sampling without replacement results on the average in an overestimation of $Var_{wor}(\hat{Y}_{HT})$. If $n = 3$, an underestimation may occur, but this underestimation has been found to be less than 5%.
If rejective sampling is applied, the use of $\hat{Var}_{wr}(\hat{Y}_{pps})$ under sampling without replacement may result on the average in an underestimation of $Var_{wor}(\hat{Y}_{HT})$, irrespective of whether $n = 2$ or $n = 3$. However if $n = 2$, this underestimation is less than 11% while if $n = 3$, this underestimation may be as large as 44%.

TABLE 3.4

VALUES OF $E_{wor}[\hat{Var}_{wor}(\hat{Y}_{pps})]/Var_{wor}(\hat{Y}_{HT})$
FOR EACH POPULATION

				SUCCESSIVE SAMPLING				*REJECTIVE SAMPLING*			
				N				*N*			
Case	*Set*	*Case*	*n*	8	10	12	15	8	10	12	15
I	1	*A*	2	1,182	1,274	1,089	1,136	0,922	1,082	0,929	1,039
			3	0,980	1,135	0,989	1,107	0,567	0,657	0,672	0,820
		B	2	1,435	1,321	1,255	1,186	1,566	1,415	1,329	1,241
			3	1,943	1,642	1,488	1,342	2,311	1,884	1,671	1,470
		C	2	1,466	1,332	1,294	1,202	1,629	1,437	1,412	1,274
			3	2,029	1,666	1,590	1,380	2,500	1,937	1,897	1,550
	2	*A*	2	1,220	1,228	1,132	1,179	1,059	1,165	1,042	1,165
			3	1,218	1,323	1,142	1,277	0,898	1,148	0,944	1,214
		B	2	1,435	1,321	1,255	1,188	1,566	1,415	1,329	1,244
			3	1,942	1,641	1,488	1,346	2,309	1,884	1,670	1,479
		C	2	1,473	1,341	1,271	1,199	1,618	1,449	1,364	1,260
			3	2,029	1,688	1,528	1,367	2,430	1,970	1,757	1,512
II	1	*A*	2	1,594	1,195	1,028	1,098	1,541	1,085	0,895	1,022
			3	1,742	1,181	0,954	1,098	1,123	0,880	0,722	0,923
		B	2	1,430	1,294	1,268	1,189	1,531	1,379	1,331	1,234
			3	1,881	1,572	1,500	1,338	2,153	1,790	1,651	1,442
		C	2	1,429	1,334	1,249	1,213	1,488	1,419	1,306	1,257
			3	1,838	1,652	1,458	1,380	1,963	1,870	1,591	1,479
	2	*A*	2	1,246	1,341	1,257	1,132	1,076	1,433	1,268	1,111
			3	1,197	1,657	1,431	1,199	0,806	1,843	1,406	1,147
		B	2	1,433	1,309	1,267	1,191	1,548	1,386	1,330	1,236
			3	1,904	1,594	1,497	1,342	2,219	1,786	1,650	1,447
		C	2	1,433	1,285	1,231	1,174	1,570	1,320	1,278	1,201
			3	1,915	1,517	1,419	1,301	2,287	1,591	1,519	1,355

In the case where the PSUs are not of equal MOS (Cases I and II) and M_i is not correlated with y_i (Cases B and C), the use of $\hat{Var}_{wr}(\hat{Y}_{pps})$ leads on the average to an overestimation of $Var_{wor}(\hat{Y}_{HT})$. By comparing Tables 3.3 and 3.4, it can be seen that the ratio given by (3.34) is always smaller than that given by (3.37). Thus in this case, $Var_{wor}(\hat{Y}_{pps})$ always exceeds $Var_{wor}(\hat{Y}_{HT})$.

The conclusions that have been drawn from Table 3.3 are also applicable in this case.

3.3.4 Comparison between $\hat{Var}_{wr}(\hat{Y}_{pps})$ and $\hat{Var}_{wor}(\hat{Y}_{pps})$

In order to compare the individual estimators $\hat{Var}_{wr}(\hat{Y}_{pps})$ and $\hat{Var}_{wor}(\hat{Y}_{pps})$, the ratio (3.38) has been calculated.

For each population, the total probability that a sample is selected for which this ratio is less than 1, i.e. for which $\hat{Var}_{wr}(\hat{Y}_{pps})$ is less than $\hat{Var}_{wor}(\hat{Y}_{pps})$, is given in Table 3.5.

By comparing the individual estimators, it has been found that if $\hat{Var}_{wr}(\hat{Y}_{pps})$ is large, $\hat{Var}_{wor}(\hat{Y}_{pps})$ is not necessarily large and vice versa. Thus the variations in $\hat{Var}_{wr}(\hat{Y}_{pps})$ and $\hat{Var}_{wor}(\hat{Y}_{pps})$ may be in opposite directions. The individual values of the ratio (3.38) may vary greatly and the maximum values, means and standard deviations of this ratio may be very large for a particular population.

In order to investigate how the mean of $\hat{Var}_{wr}(\hat{Y}_{pps})$ and $\hat{Var}_{wor}(\hat{Y}_{pps})$ differ for each population, the ratio of these two means has been calculated. These values are given in Table 3.6.
Also, the standard deviation for each estimator has been calculated and the ratios of these two standard deviations are given in Table 3.7 for each population.

From inspection of these three tables it can be seen that the results for rejective and successive sampling are very similar and that there is hardly any difference between these two sampling methods. Therefore, conclusions will only be drawn for successive sampling and they will also be true for rejective sampling.

From Table 3.5 it can be seen that for each population it is possible to select a sample for which $\hat{Var}_{wr}(\hat{Y}_{pps})$ is less than $\hat{Var}_{wor}(\hat{Y}_{pps})$. Thus, although the expected value of $\hat{Var}_{wr}(\hat{Y}_{pps})$ exceeds the expected value of $\hat{Var}_{wor}(\hat{Y}_{pps})$, the individual values of $\hat{Var}_{wr}(\hat{Y}_{pps})$ may be smaller than those of $\hat{Var}_{wor}(\hat{Y}_{pps})$. The total probability of selecting a sample for which $\hat{Var}_{wr}(\hat{Y}_{pps})$ is smaller than $\hat{Var}_{wor}(\hat{Y}_{pps})$ varies between 40% and 59% when $n = 2$ and between 20% and 49% when $n = 3$. For a specific N, this probability is always larger for $n = 2$ than for $n = 3$. The probability of selecting such a sample in a single

TABLE 3.5

TOTAL PROBABILITY OF SELECTING
A SAMPLE FOR WHICH $\hat{Var}_{wr}(\hat{Y}_{pps}) < \hat{Var}_{wor}(\hat{Y}_{pps})$
FOR EACH POPULATION

				SUCCESSIVE SAMPLING				*REJECTIVE SAMPLING*			
				N				*N*			
Case	*Set*	*Case*	*n*	8	10	12	15	8	10	12	15
I	1	*A*	2	0,418	0,439	0,455	0,461	0,419	0,440	0,456	0,461
			3	0,345	0,354	0,397	0,339	0,344	0,359	0,417	0,341
		B	2	0,491	0,505	0,515	0,547	0,484	0,500	0,510	0,543
			3	0,238	0,265	0,294	0,479	0,200	0,231	0,250	0,439
		C	2	0,573	0,505	0,515	0,588	0,565	0,500	0,510	0,584
			3	0,434	0,393	0,397	0,482	0,411	0,390	0,419	0,479
	2	*A*	2	0,433	0,459	0,455	0,469	0,435	0,460	0,456	0,469
			3	0,208	0,320	0,305	0,352	0,261	0,323	0,306	0,352
		B	2	0,491	0,505	0,515	0,547	0,484	0,500	0,510	0,543
			3	0,228	0,274	0,299	0,483	0,200	0,236	0,264	0,446
		C	2	0,573	0,566	0,515	0,461	0,565	0,560	0,510	0,461
			3	0,414	0,435	0,4017	0,427	0,400	0,415	0,378	0,431
II	1	*A*	2	0,430	0,463	0,460	0,495	0,430	0,463	0,459	0,495
			3	0,371	0,322	0,373	0,416	0,372	0,322	0,381	0,412
		B	2	0,442	0,528	0,453	0,468	0,441	0,525	0,453	0,467
			3	0,408	0,529	0,355	0,423	0,376	0,510	0,358	0,420
		C	2	0,477	0,462	0,462	0,481	0,476	0,460	0,461	0,483
			3	0,334	0,286	0,398	0,351	0,337	0,285	0,389	0,353
	2	*A*	2	0,458	0,452	0,453	0,463	0,458	0,451	0,453	0,463
			3	0,221	0,302	0,284	0,359	0,223	0,296	0,287	0,359
		B	2	0,453	0,508	0,452	0,489	0,450	0,505	0,452	0,488
			3	0,291	0,437	0,336	0,332	0,286	0,431	0,334	0,350
		C	2	0,453	0,447	0,458	0,469	0,450	0,447	0,459	0,469
			3	0,212	0,323	0,419	0,332	0,207	0,314	0,411	0,332

TABLE 3.5 (CONT.)

				SUCCESSIVE SAMPLING				*REJECTIVE SAMPLING*			
				N				*N*			
Case	*Set*	*Case*	*n*	8	10	12	15	8	10	12	15
III	1	*A*	2	0,429	0,467	0,470	0,486	0,429	0,467	0,470	0,486
			3	0,286	0,317	0,341	0,374	0,286	0,317	0,341	0,374
		B	2	0,464	0,467	0,470	0,467	0,464	0,467	0,470	0,467
			3	0,304	0,408	0,373	0,319	0,304	0,408	0,373	0,319
		C	2	0,429	0,444	0,470	0,524	0,429	0,444	0,470	0,524
			3	0,179	0,233	0,345	0,411	0,170	0,233	0,345	0,411
	2	*A*	2	0,464	0,467	0,470	0,486	0,464	0,467	0,470	0,486
			3	0,339	0,383	0,345	0,330	0,339	0,383	0,345	0,330
		B	2	0,464	0,444	0,470	0,467	0,464	0,444	0,470	0,467
			3	0,286	0,275	0,323	0,396	0,286	0,275	0,323	0,396
		C	2	0,429	0,467	0,470	0,486	0,429	0,467	0,470	0,486
			3	0,250	0,283	0,318	0,345	0,250	0,283	0,318	0,345

stratum is thus quite high. Since $\hat{Var}_{wr}(\hat{Y}_{pps})$ may sometimes be larger and sometimes be smaller than $\hat{Var}_{wor}(\hat{Y}_{pps})$ and since sampling in each stratum takes place independently, a large number of strata are needed in order for this variation to cancel out. Thus if a large number of strata are used, $\hat{Var}_{wr}(\hat{Y}_{pps})$ can safely be used as an estimate of $\hat{Var}_{wor}(\hat{Y}_{pps})$.

From Table 3.6 it can be seen that the mean of $\hat{Var}_{wr}(\hat{Y}_{pps})$ is larger than that of $\hat{Var}_{wor}(\hat{Y}_{pps})$, except if Case I is applied, M_i is correlated with $1/y_i$ (Case B) and N is large. However when this exception occurs, it can be seen that the mean of $\hat{Var}_{wr}(\hat{Y}_{pps})$ is at most 6% smaller than the mean of $\hat{Var}_{wor}(\hat{Y}_{pps})$. It can also be seen that the two mean values are closer to each other when N is larger and when $n = 2$.

The ratios of the standard deviations are larger than one when $n = 2$. However when $n = 3$, these ratios are smaller than one except if $N = 8$ and Case II is applied. Thus the values of $\hat{Var}_{wr}(\hat{Y}_{pps})$ vary more than those of $\hat{Var}_{wor}(\hat{Y}_{pps})$ when $n = 2$ and vice versa when $n = 3$ (except if $N = 8$ and Case II is applied).

It is thus advisable that a sample of size three is selected and if Case II is applied that more than 8 PSUs be formed. In this case, the probability of selecting a sample for which $\hat{Var}_{wr}(\hat{Y}_{pps})$ is less than $\hat{Var}_{wor}(\hat{Y}_{pps})$ is lower and $\hat{Var}_{wr}(\hat{Y}_{pps})$ varies less than $\hat{Var}_{wor}(\hat{Y}_{pps})$. However the mean of $\hat{Var}_{wr}(\hat{Y}_{pps})$ may exceed the mean of $\hat{Var}_{wor}(\hat{Y}_{pps})$.

TABLE 3.6

RATIOS OF THE MEAN VALUES
FOR EACH POPULATION

				SUCCESSIVE SAMPLING				*REJECTIVE SAMPLING*			
				N				*N*			
Case	*Set*	*Case*	*n*	8	10	12	15	8	10	12	15
I	1	*A*	2	1,332	1,241	1,193	1,154	1,332	1,242	1,193	1,154
			3	1,597	1,413	1,321	1,250	1,598	1,416	1,323	1,250
		B	2	1,092	1,023	0,982	0,952	1,116	1,040	0,996	0,962
			3	1,205	1,072	1,000	0,947	1,282	1,125	1,041	0,975
		C	2	1,148	0,097	1,034	1,012	1,167	1,110	1,045	1,019
			3	1,290	1,182	1,074	1,032	1,354	1,222	1,108	1,054
	2	*A*	2	1,297	1,243	1,172	1,153	1,301	1,243	1,175	1,153
			3	1,535	1,416	1,288	1,249	1,550	1,419	1,295	1,249
		B	2	1,092	1,023	0,982	0,952	1,116	1,041	0,996	0,962
			3	1,204	1,072	1,000	0,947	1,281	1,126	1,041	0,975
		C	2	1,168	1,074	1,053	1,030	1,185	1,088	1,063	1,036
			3	1,322	1,147	1,102	1,058	1,380	1,192	1,132	1,078
II	1	*A*	2	1,312	1,248	1,189	1,152	1,315	1,248	1,190	1,152
			3	1,563	1,426	1,315	1,247	1,571	1,426	1,318	1,247
		B	2	1,250	1,142	1,126	1,086	1,260	1,151	1,132	1,090
			3	1,457	1,250	1,215	1,143	1,490	1,279	1,233	1,156
		C	2	1,304	1,154	1,154	1,114	1,308	1,162	1,157	1,116
			3	1,549	1,269	1,258	1,186	1,562	1,296	1,270	1,195
	2	*A*	2	1,333	1,229	1,196	1,152	1,333	1,231	1,196	1,152
			3	1,598	1,393	1,326	1,247	1,599	1,399	1,327	1,248
		B	2	1,207	1,186	1,109	1,097	1,221	1,192	1,116	1,100
			3	1,386	1,322	1,188	1,160	1,433	1,341	1,210	1,171
		C	2	1,251	1,236	1,174	1,144	1,260	1,237	1,176	1,145
			3	1,457	1,405	1,290	1,234	1,488	1,410	1,296	1,236

TABLE 3.6 (CONT.)

				SUCCESSIVE SAMPLING				*REJECTIVE SAMPLING*			
				N				*N*			
Case	*Set*	*Case*	*n*	8	10	12	15	8	10	12	15
III	1	*A*	2	1,333	1,250	1,200	1,154	1,333	1,250	1,200	1,154
			3	1,660	1,429	1,333	1,250	1,600	1,429	1,333	1,154
		B	2	1,333	1,250	1,200	1,154	1,333	1,250	1,200	1,154
			3	1,660	1,429	1,333	1,250	1,600	1,429	1,333	1,154
		C	2	1,333	1,250	1,200	1,154	1,333	1,250	1,200	1,154
			3	1,660	1,429	1,333	1,250	1,600	1,429	1,333	1,154
	2	*A*	2	1,333	1,250	1,200	1,154	1,333	1,250	1,200	1,154
			3	1,660	1,429	1,333	1,250	1,600	1,429	1,333	1,154
		B	2	1,333	1,250	1,200	1,154	1,333	1,250	1,200	1,154
			3	1,660	1,429	1,333	1,250	1,600	1,429	1,333	1,154
		C	2	1,333	1,250	1,200	1,154	1,333	1,250	1,200	1,154
			3	1,660	1,429	1,333	1,250	1,600	1,429	1,333	1,154

Conclusions regarding the ratios of the means and the standard deviations can now be drawn. However these conclusions are often only tendencies and not very clear.

The following conclusions regarding the ratio of the means can be drawn from Table 3.6:

(i) If the PSUs are formed in such a way that half of the measures of size are equal to the lower bound of (2.37) and the other half equal to the upper bound of (2.37), i.e. Case I, the mean of $\hat{Var}_{wr}(\hat{Y}_{pps})$ is closer to that of $\hat{Var}_{wor}(\hat{Y}_{pps})$ if M_i is correlated with $1/y_i$ (Case B). The two means differ the most if M_i is correlated with y_i,i.e. Case A.

(ii) If the PSUs are formed in such a way that the M_i's are evenly spread over range (2.37), i.e. Case II, the same result as in (i) is obtained.

(iii) If the PSUs are of equal measure of size (Case III), the difference between the two means is the same regardless whether M_i is correlated with y_i or not.

(iv) If M_i is correlated with y_i (Case A), the difference between the two means is the largest if Case III is applied. However, the values of the ratio for Cases I, II and III lie very close to each other.

(v) If M_i is correlated with $1/y_i$ (Case B), the difference between the mean of $\hat{Var}_{wr}(\hat{Y}_{pps})$ and $\hat{Var}_{wor}(\hat{Y}_{pps})$ is the smallest for Case I and the largest for Case III.

TABLE 3.7

RATIOS OF THE STANDARD DEVIATIONS FOR EACH POPULATION

				SUCCESSIVE SAMPLING				*REJECTIVE SAMPLING*			
				N				*N*			
Case	*Set*	*Case*	*n*	8	10	12	15	8	10	12	15
I	1	*A*	2	1,369	1,217	1,129	1,090	1,359	1,216	1,130	1,090
			3	0,949	0,907	0,761	0,693	0,937	0,909	0,763	0,693
		B	2	1,111	1,082	1,069	1,069	1,101	1,076	1,065	1,066
			3	0,648	0,629	0,626	0,657	0,619	0,607	0,609	0,644
		C	2	1,220	1,226	1,158	1,122	1,214	1,227	1,157	1,122
			3	0,852	0,990	0,859	0,832	0,828	0,981	0,848	0,824
	2	*A*	2	1,270	1,156	1,167	1,091	1,270	1,155	1,166	1,090
			3	0,900	0,759	0,788	0,674	0,901	0,759	0,786	0,673
		B	2	1,109	1,084	1,069	1,070	1,099	1,077	1,065	1,067
			3	0,645	0,631	0,626	0,659	0,617	0,610	0,610	0,646
		C	2	1,225	1,147	1,196	1,145	1,224	1,143	1,197	1,146
			3	0,887	0,764	0,982	0,920	0,874	0,749	0,976	0,917
II	1	*A*	2	1,221	1,082	1,186	1,152	1,219	1,083	1,186	1,152
			3	0,813	0,667	0,887	0,815	0,812	0,669	0,886	0,814
		B	2	1,269	1,197	1,129	1,122	1,268	1,195	1,129	1,122
			3	0,972	0,897	0,776	0,828	0,962	0,883	0,774	0,825
		C	2	1,184	1,133	1,178	1,101	1,184	1,133	1,180	1,101
			3	0,809	0,755	0,933	0,728	0,806	0,751	0,933	0,729
	2	*A*	2	1,457	1,153	1,098	1,088	1,459	1,154	1,098	1,089
			3	1,258	0,792	0,677	0,709	1,254	0,793	0,677	0,710
		B	2	1,217	1,224	1,124	1,111	1,215	1,224	1,124	1,111
			3	0,888	0,953	0,770	0,794	0,875	0,947	0,768	0,792
		C	2	1,303	1,128	1,152	1,131	1,308	1,128	1,153	1,132
			3	1,107	0,734	0,837	0,836	1,109	0,734	0,836	0,837

TABLE 3.7 (CONT.)

				SUCCESSIVE SAMPLING				*REJECTIVE SAMPLING*			
				N				*N*			
Case	*Set*	*Case*	*n*	8	10	12	15	8	10	12	15
III	1	*A*	2	1,280	1,126	1,080	1,122	1,280	1,126	1,080	1,122
			3	0,943	0,717	0,652	0,771	0,943	0,717	0,652	0,771
		B	2	1,109	1,147	1,118	1,116	1,109	1,147	1,118	1,116
			3	0,672	0,753	0,724	0,756	0,672	0,753	0,724	0,756
		C	2	1,241	1,193	1,112	1,117	1,241	1,193	1,112	1,117
			3	0,879	0,837	0,712	0,758	0,879	0,837	0,712	0,758
	2	*A*	2	1,176	1,129	1,128	1,066	1,176	1,129	1,128	1,066
			3	0,775	0,723	0,743	0,643	0,775	0,723	0,743	0,643
		B	2	1,248	1,069	1,088	1,108	1,248	1,069	1,088	1,108
			3	0,891	0,618	0,666	0,738	0,891	0,618	0,666	0,738
		C	2	1,069	1,175	1,086	1,104	1,069	1,175	1,086	1,104
			3	0,609	0,804	0,662	0,728	0,609	0,804	0,662	0,728

(vi) If M_i is independent of y_i (Case C), the same result as in (v) is obtained.

The following conclusions regarding the ratio of the standard deviations can be drawn from Table 3.7:

(i) If the PSUs are formed in such a way that half of the measures of size are equal to the lower bound of (2.37) and the other half equal to the upper bound of (2.37), i.e. Case I, the standard deviation of $\hat{Var}_{wr}(\hat{Y}_{pps})$ is closer to that of $\hat{Var}_{wor}(\hat{Y}_{pps})$ if M_i is correlated with $1/y_i$ (Case B).

(ii) If the PSUs are formed in such a way that the M_i's are evenly spread over range (2.37), i.e. Case II, it is not clear whether the M_i's should be correlated with y_i or not, in order for the difference between the two standard deviations to be the smallest.

(iii) If the PSUs are of equal measure of size, i.e. Case III, the same result as in (ii) is obtained.

(iv) If M_i is correlated with y_i (Case A), it is not clear whether Cases I, II or III should be applied in order for the difference between the standard deviations of $\hat{Var}_{wr}(\hat{Y}_{pps})$ and $\hat{Var}_{wor}(\hat{Y}_{pps})$ to be the smallest.

(v) If M_i is correlated with $1/y_i$ (Case B), the difference between the standard deviation of $\hat{Var}_{wr}(\hat{Y}_{pps})$ and $\hat{Var}_{wor}(\hat{Y}_{pps})$ is the smallest for Case I and the largest for Case II.

(vi) If M_i is independent of y_i (Case C), the difference between the standard deviation of $\hat{Var}_{wr}(\hat{Y}_{pps})$ and $\hat{Var}_{wor}(\hat{Y}_{pps})$ is the largest if the PSUs are formed according to Case I.

In order for the difference between the mean values and standard deviations of $\hat{Var}_{wr}(\hat{Y}_{pps})$ and $\hat{Var}_{wor}(\hat{Y}_{pps})$ to be the smallest, the following recommendations can thus be made:

(i) If the PSUs are not of equal measure of size, M_i should be correlated with $1/y_i$.

(ii) If the PSUs are of equal measure of size, it does not matter whether M_i is correlated with y_i or not.

(iii) If M_i is correlated with y_i, the PSUs can be formed in any way, as long as the M_i's lie within the limits (2.37).

(iv) If M_i is correlated with $1/y_i$, the PSUs should be formed according to Case I, i.e. the variation between the M_i's should be large.

(v) If M_i is independent of y_i, the PSUs should be formed according to Case II.

(vi) If there exists any doubt whether M_i is correlated with y_i or not, it is thus advisable that the PSUs are of equal measure of size.

CHAPTER 4

CONCLUSIONS AND RECOMMENDATIONS

In Chapters 2 and 3 it has been numerically verified that the Kish assertion is true on the average, regardless whether the inclusion probabilities are equal to nz_i or not. This means, the use of $\hat{Var}_{wr}(\hat{Y}_{pps})$ instead of $\hat{Var}_{wor}(\hat{Y}_{pps})$ under sampling without replacement leads on the average to an overestimation of $Var_{wor}(\hat{Y}_{pps})$.

In the case where $\pi_i = nz_i$, the two population total estimators $\hat{Y}_{pps}$ and $\hat{Y}_{HT}$ are identical and $\hat{Y}_{pps}$ can be used without any bias. If $\pi_i \neq nz_i$, the two estimators are not identical and the use of $\hat{Y}_{pps}$ will lead to a biased estimation of the population total. However it has been found that this bias is negligible and will furthermore tend to cancel out when several strata are used.

By comparing the individual estimators $\hat{Var}_{wr}(\hat{Y}_{pps})$ and $\hat{Var}_{wor}(\hat{Y}_{pps})$, it has been found that $\hat{Var}_{wr}(\hat{Y}_{pps})$ exceeds $\hat{Var}_{wor}(\hat{Y}_{pps})$ if $\pi_i = nz_i$. Since both estimators are subject to variation, both may over- or underestimate $Var_{wor}(\hat{Y}_{pps})$. Since sampling takes place independently in each stratum, it is expected that this variation in over- and underestimation will tend to cancel out if several strata are used and the probability that the use of $\hat{Var}_{wr}(\hat{Y}_{pps})$ will lead to an underestimation of $Var_{wor}(\hat{Y}_{pps})$ will become very small.
If the sampling scheme applied does not ensure that $\pi_i = nz_i$, it has been found that $\hat{Var}_{wr}(\hat{Y}_{pps})$ need not necessarily exceed $\hat{Var}_{wor}(\hat{Y}_{pps})$. It may happpen that if $\hat{Var}_{wor}(\hat{Y}_{pps})$ is large, $\hat{Var}_{wor}(\hat{Y}_{pps})$ is small and vice versa. The probability that a sample is selected for which $\hat{Var}_{wr}(\hat{Y}_{pps})$ is smaller than $\hat{Var}_{wor}(\hat{Y}_{pps})$ is quite large, but less than 0,60 when $n = 2$ and 0,50 when $n = 3$. Thus also in this case, a large number of strata are needed in order to ensure that $\hat{Var}_{wr}(\hat{Y}_{pps})$ can safely be used to estimate $Var_{wr}(\hat{Y}_{pps})$ under sampling without replacement.

It therefore seems that a sampling procedure for which $\pi_i = nz_i$ is advisable.

In Chapter 2 it has been found that Durbin's procedure and Sampford's rejective procedure are the easiest to apply in practice when $n = 2$ and $n = 3$, respectively. Since Sampford's rejective procedure can also be applied when a sample of size 2 is selected, it is recommended that Sampford's rejective procedure be used in practice.

If this procedure is applied, it is advisable that a stratum comprises more than 8 PSUs and that a sample of size 2 is selected from these PSUs. It is also recommendable that the PSUs are of equal measure of size, but since this is impracticable, it is recommended that the measures of size are evenly spread over the range (2.37).

APPENDIX A

POPULATIONS USED IN THE NUMERICAL INVESTIGATIONS

N = 8

Set										
					Case I					
		M_i	250	250	250	250	500	500	500	500
	y_i	Case A	80	85	95	92	172	178	198	163
		Case B	139	128	144	148	67	76	69	71
		Case C	179	95	157	170	93	151	172	126
					Case II					
		M_i	475	326	348	333	279	432	418	268
	y_i	Case A	174	114	127	129	109	164	143	104
1		Case B	68	100	93	95	128	75	85	132
		Case C	116	165	109	101	88	85	180	136
					Case III					
		M_i	500	500	500	500	500	500	500	500
	y_i	Case A	173	180	187	171	186	160	169	197
		Case B	68	73	74	75	68	74	72	67
		Case C	131	164	174	109	135	162	138	167
					Case I					
		M_i	2500	2500	2500	2500	5000	5000	5000	5000
	y_i	Case A	535	493	438	398	1111	915	1227	864
		Case B	470	493	437	498	232	240	249	234
		Case C	1228	1010	1301	694	730	1329	692	1410
					Case II					
		M_i	4670	2579	3738	3209	4861	4389	2534	3103
	y_i	Case A	875	480	629	645	991	893	527	633
2		Case B	707	1338	888	1082	695	784	1286	1044
		Case C	1071	876	625	1103	1216	896	1257	1165
					Case III					
		M_i	5000	5000	5000	5000	5000	5000	5000	5000
	y_i	Case A	824	949	926	777	962	1056	804	1004
		Case B	223	241	251	227	222	233	229	242
		Case C	522	628	1439	1411	1371	502	873	1255

Set											
		Case I									
		M_i	250	250	250	250	250	500	500	500	500	500
	y_i	Case A	93	91	82	95	100	192	166	179	181	185
		Case B	150	147	144	135	140	69	75	74	70	67
		Case C	170	113	180	96	102	83	169	151	93	110
		Case II										
		M_i	389	435	369	471	451	488	250	381	272	266
1	y_i	Case A	148	154	129	182	157	164	81	148	106	92
		Case B	113	109	125	90	100	89	166	107	165	170
		Case C	108	149	154	128	93	91	139	172	162	131
		Case III										
		M_i	500	500	500	500	500	500	500	500	500	500
	y_i	Case A	187	167	194	178	200	180	168	170	198	160
		Case B	74	69	73	69	68	70	67	69	74	71
		Case C	172	123	126	171	162	169	115	154	132	88
		Case I										
		M_i	2500	2500	2500	2500	2500	5000	5000	5000	5000	5000
	y_i	Case A	396	620	435	498	380	1106	947	1234	789	858
		Case B	447	500	475	493	460	218	246	228	239	252
		Case C	1178	1403	1346	721	1327	934	0599	1482	941	1199
		Case II										
		M_i	3708	3460	4423	4016	3163	2728	4663	3516	3875	2705
2	y_i	Case A	595	783	787	748	648	652	858	538	815	576
		Case B	1069	1200	980	1001	1362	1504	858	1134	1093	1593
		Case C	1406	756	713	896	544	916	1303	1410	1266	1215
		Case III										
		M_i	5000	5000	5000	5000	5000	5000	5000	5000	5000	5000
	y_i	Case A	1229	960	1216	1029	914	1236	850	997	865	918
		Case B	251	222	219	235	226	217	243	250	250	216
		Case C	981	1201	528	830	574	812	1271	1099	1086	1163

Set													
		Case I											
		M_i	250	250	250	250	250	250	500	500	500	500	500	500
	y_i	Case A	100	92	88	90	86	80	184	168	200	175	173	196
		Case B	145	141	132	152	137	143	71	74	72	74	68	73
		Case C	180	168	83	153	123	126	107	143	108	104	132	138
		Case II												
		M_i	378	494	322	312	498	411	356	279	455	309	344	424
1	y_i	Case A	133	192	107	123	181	146	134	98	173	111	110	156
		Case B	117	89	136	134	82	105	120	147	99	134	120	104
		Case C	129	132	108	133	114	177	103	170	98	153	111	97
		Case III												
		M_i	500	500	500	500	500	500	500	500	500	500	500	500
	y_i	Case A	166	185	186	162	171	194	165	198	174	196	187	198
		Case B	69	76	69	70	72	71	72	68	76	74	72	75
		Case C	100	163	146	163	107	135	177	139	96	117	146	87
		Case I												
		M_i	2500	2500	2500	2500	2500	2500	5000	5000	5000	5000	5000	5000
	y_i	Case A	533	442	471	393	479	420	909	984	1086	1114	1243	861
		Case B	469	481	499	440	439	474	251	237	231	227	229	248
		Case C	1321	843	817	1023	828	1485	993	1313	1322	855	1164	687
		Case II												
		M_i	4893	4604	4055	2719	4787	3435	2920	3135	2601	3582	3924	3237
2	y_i	Case A	794	1057	855	513	809	726	662	573	530	747	665	505
		Case B	811	939	1068	1489	912	1262	1365	1393	1555	1207	1043	1330
		Case C	1331	1282	699	997	1406	1001	926	821	1137	617	751	1230
		Case III												
		M_i	5000	5000	5000	5000	5000	5000	5000	5000	5000	5000	5000	5000
	y_i	Case A	833	1106	967	928	1239	755	811	764	787	859	1068	1166
		Case B	250	216	223	241	219	249	221	235	237	239	224	244
		Case C	1452	1371	788	1309	1273	943	762	1395	573	869	579	1010

N=15

Set									
		Case I							
		M_i	250	250	250	250	250	250	250	
	y_i	Case A	81	87	90	97	95	88	85	
		Case B	145	152	147	134	141	139	150	
		Case C	176	84	172	159	104	161	120	
		M_i	500	500	500	500	500	500	500	500
	y_i	Case A	184	162	192	176	188	183	166	170
		Case B	71	72	77	70	67	64	68	65
		Case C	171	150	119	176	123	101	114	166
		Case II								
		M_i	425	293	346	388	467	266	344	
	y_i	Case A	137	104	127	154	170	101	128	
		Case B	91	146	123	111	84	159	117	
		Case C	121	126	168	152	108	127	134	
1		M_i	256	353	295	426	368	313	468	404
	y_i	Case A	93	119	106	161	127	102	184	147
		Case B	162	120	137	91	105	134	91	100
		Case C	98	162	164	125	168	122	145	101
		Case III								
		M_i	500	500	500	500	500	500	500	
	y_i	Case A	193	172	173	184	199	178	168	
		Case B	74	74	71	67	70	77	71	
		Case C	146	125	144	110	94	98	141	
		M_i	500	500	500	500	500	500	500	500
	y_i	Case A	164	176	161	176	169	183	200	187
		Case B	72	76	71	67	74	77	71	76
		Case C	172	175	139	162	90	139	144	153

Set									
		Case I							
		M_i	2500	2500	2500	2500	2500	2500	2500	
	y_i	Case A	475	586	564	499	556	376	431	
		Case B	461	434	476	502	469	443	459	
		Case C	1110	835	938	735	1397	1429	715	
		M_i	5000	5000	5000	5000	5000	5000	5000	5000
	y_i	Case A	929	1204	950	797	1108	862	785	1229
		Case B	241	244	237	224	250	228	244	249
		Case C	1422	1060	1071	919	1320	704	1385	553
		Case II								
		M_i	3080	3367	2584	2978	3884	2784	3044	
	y_i	Case A	552	519	467	464	801	673	663	
		Case B	1395	1246	1568	1425	1088	1421	1375	
		Case C	1026	726	956	1319	1270	782	630	
2		M_i	4173	3832	4056	3925	4560	3305	3793	4412
	y_i	Case A	676	884	890	717	1025	636	706	892
		Case B	966	1060	989	1085	919	1250	1058	903
		Case C	1111	1192	1073	858	1447	1197	930	908
		Case III								
		M_i	5000	5000	5000	5000	5000	5000	5000	
	y_i	Case A	865	1099	1223	819	1162	1113	846	
		Case B	229	246	216	245	234	235	239	
		Case C	1260	1473	540	586	618	697	890	
		M_i	5000	5000	5000	5000	5000	5000	5000	5000
	y_i	Case A	870	986	1211	935	1192	758	761	951
		Case B	223	248	243	218	239	225	251	238
		Case C	518	1065	1493	1223	806	712	1083	879

APPENDIX B

TABLE B.1

VALUES OF $bias^2/Var_{wor}(\hat{Y}_{pps})$ FOR EACH POPULATION FOR SUCCESSIVE SAMPLING

				N			
Case	*Set*	*Case*	*n*	8	10	12	15
I	1	*A*	2	0.00000691	0.000027512	0.000013105	0.000000082
			3	0.00006040	0.000210390	0.000096970	0.000000580
		B	2	0.00147332	0.000822150	0.000520426	0.000275045
			3	0.01208200	0.006187750	0.003730120	0.001881570
		C	2	0.00111207	0.000538309	0.000391435	0.000188479
			3	0.00905490	0.004031800	0.002794490	0.001287140
	2	*A*	2	0.00016813	0.000020422	0.000050179	0.000000427
			3	0.00148890	0.000159280	0.000371550	0.000003030
		B	2	0.00147528	0.000821055	0.000520059	0.000274898
			3	0.01209820	0.006179290	0.003727470	0.001880450
		C	2	0.00095554	0.000616855	0.000343498	0.000159281
			3	0.00787710	0.004645140	0.002452830	0.001091930
II	1	*A*	2	0.00009460	0.000000076	0.000033989	0.000004691
			3	0.00077580	0.000000960	0.000239120	0.000032070
		B	2	0.00052934	0.000331299	0.000185402	0.000095176
			3	0.00420870	0.002430990	0.001320450	0.000647330
		C	2	0.00019758	0.000305104	0.000109291	0.000062205
			3	0.00159310	0.002282080	0.000772670	0.000427600
	2	*A*	2	0.00000001	0.000060249	0.000012647	0.000003394
			3	0.00000000	0.000447560	0.000090530	0.000022980
		B	2	0.00078072	0.000220132	0.000229922	0.000077531
			3	0.00623870	0.001619820	0.001637060	0.000527850
		C	2	0.00044170	0.000044945	0.000039261	0.000011318
			3	0.00350870	0.000333850	0.000281900	0.000077390

TABLE B.1 (CONT.)

				N			
Case	*Set*	*Case*	*n*	8	10	12	15
III	1	*A*	2	0.00000000	0.000000000	0.000000000	0.000000000
			3	0.00000000	0.000000000	0.000000000	0.000000000
		B	2	0.00000000	0.000000000	0.000000000	0.000000000
			3	0.00000000	0.000000000	0.000000000	0.000000000
		C	2	0.00000000	0.000000000	0.000000000	0.000000000
			3	0.00000000	0.000000000	0.000000000	0.000000000
	2	*A*	2	0.00000000	0.000000000	0.000000000	0.000000000
			3	0.00000000	0.000000000	0.000000000	0.000000000
		B	2	0.00000000	0.000000000	0.000000000	0.000000000
			3	0.00000000	0.000000000	0.000000000	0.000000000
		C	2	0.00000000	0.000000000	0.000000000	0.000000000
			3	0.00000000	0.000000000	0.000000000	0.000000000

TABLE B.2

VALUES OF $bias^2/Var_{wor}(\hat{Y}_{pps})$ FOR EACH POPULATION FOR REJECTIVE SAMPLING

				N			
Case	*Set*	*Case*	n	8	10	12	15
I	1	*A*	2	0.00002173	0.00009056	0.00004480	0.000000290
			3	0.00017110	0.00064290	0.00031370	0.000001970
		B	2	0.00458838	0.00269835	0.00176757	0.000970519
			3	0.03415180	0.01891680	0.01197590	0.006377040
		C	2	0.00345965	0.00176521	0.00132838	0.000664796
			3	0.02551480	0.01228910	0.00895290	0.004356780
	2	*A*	2	0.00052950	0.00006747	0.00017158	0.000001518
			3	0.00423590	0.00048940	0.00120260	0.000010390
		B	2	0.00459447	0.00269474	0.00176632	0.000969988
			3	0.03419810	0.01889060	0.01196740	0.006373120
		C	2	0.00297822	0.00202479	0.00116576	0.000562319
			3	0.02222470	0.01418570	0.00785660	0.003701110
II	1	*A*	2	0.00029821	0.00000064	0.00011158	0.000015835
			3	0.00223320	0.00000910	0.00072470	0.000100850
		B	2	0.00161119	0.00109217	0.00061907	0.000331890
			3	0.01147260	0.00748290	0.00411550	0.002148320
		C	2	0.00059311	0.00099834	0.00036637	0.000214812
			3	0.00424290	0.00693620	0.00242400	0.001396950
	2	*A*	2	0.00000008	0.00019824	0.00004117	0.000011544
			3	0.00000510	0.00137330	0.00027110	0.000073190
		B	2	0.00239245	0.00071494	0.00076775	0.000270867
			3	0.01718740	0.00486290	0.00510360	0.001756930
		C	2	0.00137641	0.00014566	0.00013704	0.000039826
			3	0.00993540	0.00099910	0.00094380	0.000260610

TABLE B.2 (CONT.)

				N			
Case	*Set*	*Case*	*n*	8	10	12	15
III	1	*A*	2	0.00000000	0.00000000	0.00000000	0.000000000
			3	0.00000000	0.00000000	0.00000000	0.000000000
		B	2	0.00000000	0.00000000	0.00000000	0.000000000
			3	0.00000000	0.00000000	0.00000000	0.000000000
		C	2	0.00000000	0.00000000	0.00000000	0.000000000
			3	0.00000000	0.00000000	0.00000000	0.000000000
	2	*A*	2	0.00000000	0.00000000	0.00000000	0.000000000
			3	0.00000000	0.00000000	0.00000000	0.000000000
		B	2	0.00000000	0.00000000	0.00000000	0.000000000
			3	0.00000000	0.00000000	0.00000000	0.000000000
		C	2	0.00000000	0.00000000	0.00000000	0.000000000
			3	0.00000000	0.00000000	0.00000000	0.000000000

REFERENCES

Brewer, K.R.W. (1963a). A model of systematic sampling with unequal probabilities. *Austr. J. Statist.* 5, 5-13.

Brewer, K.R.W. and Hanif, M. (1970). Durbin's new multistage variance estimator. *J. Roy. Statist. Soc.* Ser. B 32, 302-311.

Brewer, K.R.W. and Hanif, M. (1983). *Sampling with unequal probabilities*, Lecture Notes in Statistics, Vol. 15. Springer-Verlag, New York.

Cochran. W.G. (1977). *Sampling Techniques.* 3rd edition, John Wiley, New York.

Durbin, J. (1953). Some results in sampling theory when the units are selected with unequal probabilities. *J. Roy. Statist. Soc.*B 15, 262-269.

Durbin, J. (1967). Design of multi-stage surveys for the estimation of sampling errors. *Applied Statist.* 16, 152-164.

Gabler, S. (1981). A comparison of Sampford's sampling procedure versus unequal probability sampling with replacement. *Biometrika* 68, 725-727.

Hajek, J. (1981). *Sampling from a Finite Population.* Marcel Dekker, New York.

Horvitz, D.G. and Thompson, D.J. (1952). A generalization of sampling without replacement from a finite universe. *J. Amer. Statist. Assoc.* 47, 663-685.

Jessen, R.J. (1978). *Statistical Survey Techniques.* John Wiley, New York.

Kish, L. (1965). *Survey Sampling.* John Wiley, New York.

Konijn, H.S. (1973). *Statistical Theory of Sample Survey Design and Analysis.* North-Holland, London.

Raj, D. (1964). The use of systematic sampling with probability proportional to size in a large-scale survey. *J. Amer. Statist. Assoc.* 59, 251-255.

Raj, D. (1966). Some remarks on a simple procedure of sampling without replacement. *J. Amer. Statist. Assoc.* 61, 391-396.

Rao, J.N.K. (1965). On two simple schemes of unequal probability sampling without replacement. *J. Indian. Statist. Assoc.* 9, 173-180.

Rao, J.N.K. (1975). Unbiased variance estimation for multistage sampling designs. *Sankhyā Ser.* C 37 133-139.

Rao, J.N.K. (1988). Variance estimation in sample surveys. In *Handbook of Statistics, Vol.* 6. (P.R. Krishnaiah and C.R. Rao, eds.). North-Holland, Amsterdam, 427-447.

Sampford, M.R. (1967). On sampling without replacement with unequal probabilities of selection. *Biometrika* 45, 499-513.

Sen, A.R. (1953). On the estimate of the variance in sampling with varying probabilities. *J. Indian Soc. Agric. Statist.* 5, 119-127.

Stoker, D.J. (1984). Sampling in practice. Occasional Paper No. 11, HSRC, Pretoria.

Stoker, D.J. (1988). The analysis of complex sample data. Report WS-40, HSRC, Pretoria.

Sunter, A.B. (1986). Solutions to the problem of unequal probability sampling without replacement. *Int. Statist. Rev.* 54, 33-50.

Sutradhar, B.C. (1984). Durbin's sampling scheme versus unequal probability sampling with replacement for n=2. *Sankhyā Ser.* B 46, 370-371.

Vijayan, K. (1975). On estimating the variance in unequal probability sampling. *J. Amer. Statist. Assoc.* 70, 713-716.

Yates, F. and Grundy, P.M. (1953). Selection without replacement from within strata with probability proportional to size. *J. Roy. Statist. Soc.* B 15, 253-261.

HUMAN SCIENCES RESEARCH COUNCIL
RAAD VIR GEESTESWETENSKAPLIKE NAVORSING

Dr T van der Walt
President

Dr H C Marais
Deputy President
Adjunk-president

Dr K F Mauer
Vice-President
Vise-president

Mr/Mnr E E Garner
Man.: Finance and Corporate Services
Best.: Finansies en Korporatiewe Dienste

The HSRC endeavours to contribute, through all its functions, towards improving the quality of life of the inhabitants of South Africa

Functions of the HSRC

- determining research priorities
- undertaking, promoting, supporting and co-ordinating human sciences research
- disseminating research findings
- evaluating and promoting the implementation of research findings

Research Institutes

- Communication (ICOMM)
- Education (IER)
- History (IHR)
- Labour Economics (ILER)
- Language and the Arts (IRLA)
- Personnel (NIPR)
- Psychology and Edumetrics (IPER)
- Research Development (IRD)
- Sociology and Demography (ISODEM)
- Statistics (ISR)

Other

- Bureau for Liaison and Marketing (BLM)
- Bureau for Support Services (BSS)
- Group for Methodology and Technology (GMT)

Head office

Private Bag X41, Pretoria 0001
Republic of South Africa
Telegrams RAGEN
Telephone (012) 202 9111
Telex 32 1710 SA
Fax (012) 265362

Other addresses

National Institute for Personnel Research (NIPR)

P.O. Box 32410, Braamfontein 2017
Republic of South Africa
Telegrams NAVORSPERS
Telephone (011) 339 4451
Telex 4 25459 SA
Fax (011) 403 2353

Western Cape Regional Office

Private Bag X5, Roggebaai 8012
Tel. (021) 419 2572/3/4/5
Fax (021) 4196766

Natal Regional Office

P.O. Box 17302, Congella 4013
Tel. (031) 815 970
Fax (031) 812509

Eastern Cape Regional Office

P.O. Box 1124, Port Elizabeth 6000
Tel. (041) 532113
Telex 2 45183 SA
Fax (041) 532325

Die RGN poog om deur al sy funksies 'n bydrae te maak om die lewensgehalte van die inwoners van Suid-Afrika te verhoog

Funksies van die RGN

- bepaling van navorsingsprioriteite
- uitvoering, bevordering, ondersteuning en koördinering van geesteswetenskaplike navorsing
- disseminasie van navorsingsbevindinge
- evaluering en bevordering van die implementering van navorsingsbevindinge

Navorsingsinstitute

- Kommunikasie (IKOMM)
- Opvoedkunde (ION)
- Geskiedenis (IGN)
- Arbeidsekonomie (INAN)
- Taal en Kunste (INTAK)
- Personeel (NIPN)
- Psigologie en Edumetrika (IPEN)
- Navorsingsontwikkeling (INO)
- Sosiologie en Demografie (ISODEM)
- Statistiek (ISN)

Ander

- Buro vir Skakeling en Bemarking (BSB)
- Buro vir Ondersteunende Dienste (BOD)
- Groep vir Metodologie en Tegnologie (GMT)

Hoofkantoor

Privaatsak X41, Pretoria 0001
Republiek van Suid-Afrika
Telegramme RAGEN
Telefoon (012) 202 9111
Teleks 32 1710 SA
Faks (012) 265362

Ander adresse

Nasionale Instituut vir Personeelnavorsing (NIPN)

Posbus 32410, Braamfontein 2017
Republiek van Suid-Afrika
Telegramme NAVORSPERS
Telefoon (011) 339 4451
Teleks 4 25459 SA
Faks (011) 403 2353

Wes-Kaaplandse Streekkantoor

Privaatsak X5, Roggebaai 8012
Tel. (021) 419 2572/3/4/5
Faks (021) 4196766

Natalse Streekkantoor

Posbus 17302, Congella 4013
Tel. (031) 815 970
Faks (031) 812509

Oos-Kaaplandse Streekkantoor

Posbus 1124, Port Elizabeth 6000
Tel. (041) 532113
Teleks 2 45183 SA
Faks (041) 532325

BEST COPY AVAILABLE